TRANSACTIONS
OF THE
AMERICAN INSTITUTE
OF
ELECTRICAL ENGINEERS

A
E E
I

VOLUME 71
1952

PART I. COMMUNICATION AND ELECTRONICS

PUBLISHED BY THE
AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS
33 WEST 39TH STREET
NEW YORK 18, N. Y., U. S. A.

Copyright 1952 by the American Institute of Electrical Engineers
Printed in U. S. A. by The Mack Printing Company, Easton, Pa.

Preface

The AIEE *Transactions* for 1952 (volume 71) is published in three parts: Part I. Communication and Electronics; Part II. Applications and Industry; and Part III. Power Apparatus and Systems. The papers in each of the three parts are classified according to subject matter as follows:

Part I. Communication and Electronics

Communication Switching Systems
Radio Communications Systems
Special Communications Applications
Telegraph Systems
Television and Aural Broadcasting Systems
Wire Communications Systems
Basic Sciences
Computing Devices
Electrical Techniques in Medicine and Biology
Electronic Power Converters
Electronics
Instruments and Measurements
Magnetic Amplifiers
Metallic Rectifiers
Nucleonics

Part II. Applications and Industry

Air Transportation
Domestic and Commercial Applications
Land Transportation
Marine Transportation
Production and Application of Light
Chemical, Electrochemical and Electrothermal Applications
Electric Heating
Electric Welding
Feedback Control Systems
General Industry Applications
Industrial Control
Industrial Power Systems
Mining and Metal Industry

Part III. Power Apparatus and Systems

Carrier Current
Insulated Conductors
Power Generation
Protective Devices
Relays
Rotating Machinery
Substations
Switchgear
System Engineering
Transformers
Transmission and Distribution

Education
Safety
Research

Each part has been indexed separately in the back of that particular part. The three parts are not cross-referenced, hence the user should determine first whether the subject matter of the paper desired is in the field of communication and electronics, applications and industry, or power apparatus and systems. Papers are listed in the subject index under several key words in the titles. The original numbers assigned to the papers are given in the author index.

Volume 71 contains the technical papers and related discussions presented at these meetings:

1. Winter General Meeting, New York, N. Y., January 21–25, 1952.
2. South West District Meeting, St Louis, Mo., April 15–17, 1952.
3. North Eastern District Meeting, Binghamton, N. Y., April 30–May 2, 1952.
4. Summer General Meeting, Minneapolis, Minn., June 23–27, 1952.
5. Pacific General Meeting, Phoenix, Ariz., August 19–22, 1952.
6. AIEE Participation in Centennial of Engineering, Chicago, Ill., September 10–12, 1952.
7. Fall General Meeting, New Orleans, La., October 13–17, 1952.
8. Middle Eastern District Meeting, Toledo, Ohio, October 28–30, 1952.

The papers presented at the first three meetings were issued as AIEE *Proceedings*. These papers have been correlated according to subject matter, and combined with the papers presented at the remaining meetings which were published in the three bimonthly publications inaugurated in July and August, comprise Parts I, II, and III of volume 71 of the 1952 *Transactions*. Ultimately each part of the *Transactions* will contain the same papers and discussions that will have been published in the six successive issues of the corresponding bimonthly publication.

Statements and opinions given in the papers and discussions published in *Transactions* are the expressions of the contributors for which the American Institute of Electrical Engineers assumes no responsibility.

Communication and Electronics Index for 1952

1. Technical Subject Index

Accuracy of Single-Phase 3-Wire Metering, The Fundamental. Wentz, Petzinger.....295–302; disc. 302
Air Resistance, The Variation at Constant Density of the Dielectric Breakdown of Paper with. Cloke, Khandelwal..................309–12; disc. 312
Alloys with Rectangular Hysteresis Loops, Ultrathin Tapes of Magnetic. Littmann............220–3
A-C Instruments, Overload Protection of. Pritchett, Trau..................................13–17
A-C Network Analyzer, A Compact Inexpensive. Kimbark, Starr, Van Ness...........122–7; disc. 127
A-C Network Analyzers, A New Principle is Employed for 60-Cycle. Phillips............18–24; disc. 24
A-C Watthour Meters, Methods and Apparatus Employed in the Testing of Single-Phase. Holtz..............................61–9; disc. 69
Amplifier, A New Type of Magnetic Servo. Geyger.................................272–80
Amplifier, An Improved Magnetic Servo. Lufcy, Schmid, Barnhart.......................281–9
Amplifier as a Computer Element, The Single-Core Magnetic. Ramey.....................442–6
Amplifier Circuits, The Effect of Core Materials on Magnetic. Johnson................26–30; disc. 30
Amplifier with Inductive Loading, Series-Connected Magnetic. Wilson......................101–10
Amplifiers, Compensating for the Quiescent Current in Multistage Magnetic. FitzGerald...........8–12
Amplifiers of the Self-Balancing Potentiometer Type, Magnetic. Geyger.....................383–95
Amplifiers—Survey, Types of Magnetic. Miles..229–38
Amplifiers, The Effective Feedback Ratio of Magnetic. Finzi, Pittman, Durand..................157–64
Amplifiers, The Figure of Merit of Magnetic. Carleton, Horton..............................239–45
Analogies of a Separately Excited D-C Machine, Electromechanical. Vowels, Forte.....32–41; disc. 41
Analyzer, A Compact, Inexpensive A-C Network. Kimbark, Starr, Van Ness.....122–7; disc. 127
Analyzer Area, Lighting a Network. Dzwonczyk..............................140–6; disc. 146
Analyzer for Laplace, Poisson, Diffusion, and Wave Equations, DINA, a Digital. Leondes, Rubinoff.............................303–08; disc. 308
Analyzers, A New Principle is Employed for 60-Cycle A-C Network. Phillips...........18–24; disc. 24
Analyzers, Design Improvements and Operating Experience with 10-Kc Network. Ryder, Boast.............................437–42; disc. 442
Apparatus Employed in the Testing of Single-Phase A-C Watthour Meters, Methods and. Holtz..............................61–9; disc. 69
Automatic Regulation of Metallic Rectifiers by Magnetic Control. Smith..................111–14
Automatic Switching for Nation-wide Telephone Service. Clark, Osborne....................245–8
Automatic Toll Switching Systems. Shipley....261–9

B

(Basic Sciences) A New Property of Two-Dimensional Fields. Moore.................343–6; disc. 346
(Basic Sciences) Design of Low-Frequency Constant Time Delay Lines. Wallis..............135–40
(Basic Sciences) Design of Unequal-Q Double-Tuned Transformers. Deutsch................314–20
(Basic Sciences) Electrical and Physical Properties of IN-420: A New Chlorinated Liquid Dielectric. Warner..............................330–5
(Basic Sciences) Electromechanical Analogies of a Separately Excited D-C Machine. Vowels, Forte..............................32–41; disc. 41
(Basic Sciences) Impulse Dielectric Strength Characteristics of Liquid-Impregnated Pressboard. Dakin, Works.........................321–5; disc. 325
(Basic Sciences) Mapping Techniques Applied to Fluid Mapper Patterns. Moore...........1–5; disc. 5
(Basic Sciences) Notes on the Design of Eccles-Jordan Flip-Flops. Rubinoff...................215–20
(Basic Sciences) Solution of Electrical Engineering Problems by Southwell's Relaxation Method. Grad..............................205–14
(Basic Sciences) Subharmonics in a Series Nonlinear Circuit as Influenced by Initial Capacitor Charge. McKune, Brust.......................200–05
(Basic Sciences) Synthesis of the Transfer Function of 2-Terminal Pair Networks. Kahal.........129–34
(Basic Sciences) The Calculation of the Magnetizing Force. Halacsy.........................90–5
(Basic Sciences) The Current Status of Dynamic Stability Theory. Bothwell...................223–8
(Basic Sciences) The Derivation of Vector Potential from Tables for Scalar Potential. Smith...169–74
(Basic Sciences) The Variation at Constant Density of the Dielectric Breakdown of Paper with Air Resistance. Cloke, Khandelwal...309–12; disc. 312
(Basic Sciences) Wave Filter Characteristics by a Direct Method. Taylor, Watts..................96–100
Block Diagrams for a Transistor Digital Computer, Typical. Felker........................175–82
Breakdown of Paper with Air Resistance, The Variation at Constant Density of the Dielectric. Cloke, Khandelwal..................309–12; disc. 312
Broadcasting and Television Arts, Historic Highlights in Developing the Radio. Hanson........364–71

C

Calculation of the Magnetizing Force, The. Halacsy..............................90–5
Capacitor Charge, Subharmonics in a Series Nonlinear Circuit as Influenced by Initial. McKune, Brust..............................200–05
Carrier Telephone, Type-O. Coy, Van Tassel..428–37
Cathode Tubes, Characteristic Shifts in Oxide. Bartley, White..............................43–9
Century, The Impress of Enterprise Upon Telegraphy's First. Coggeshall.............375–81; disc. 381
Characteristic Impedance of Rectangular Coaxial Transmission Lines. Omar, Miller.........81–9
Characteristic Shifts in Oxide Cathode Tubes. Bartley, White..............................43–9
Characteristics by a Direct Method, Wave Filter. Taylor, Watts..............................96–100
Characteristics of Liquid-Impregnated Pressboard, Impulse Dielectric Strength. Dakin, Works..............................321–5; disc. 325
(Chart) Types of Magnetic Amplifiers—Survey. Miles..............................229–38
Chlorinated Liquid Dielectric, Electrical and Physical Properties of IN-420: A New. Warner....330–5
Circuit as Influenced by Initial Capacitor Charge, Subharmonics in a Series Nonlinear. McKune, Brust..............................200–05
Circuits and Devices for Industrial Process Functions, Computing. Hornfeck..................183–93
Circuits, The Effect of Core Materials on Magnetic Amplifier. Johnson..............26–30; disc. 30
Coaxial Transmission Lines, Characteristic Impedance of Rectangular. Omar, Miller............81–9
(Communication) A New General-Purpose Relay for Telephone Switching Systems. Keller.....413–28
Communication Arts, A Milestone in the. Bramhall..............................347
(Communication) Automatic Switching for Nationwide Telephone Service. Clark, Osborne........245–8
(Communication) Automatic Toll Switching Systems. Shipley..............................261–9
(Communication) Fundamental Plans for Toll Telephone Plant. Pilliod......................248–56
(Communication) Historic Highlights in Developing the Radio Broadcasting and Television Arts. Hanson..............................364–71
(Communication) Nationwide Numbering Plan. Nunn..............................257–60
(Communication) The Development of Telephony in the United States. Clark...................348–64
Communication, The First 50 Years of International Radio. Pratt..............................371–5
(Communication) Type-O Carrier Telephone. Coy, Van Tassel..............................428–37
Compensating for the Quiescent Current in Multistage Magnetic Amplifiers. FitzGerald...........7–12
Computer Element, The Single-Core Magnetic Amplifier as a. Ramey..............................442–6
Computer, Typical Block Diagrams for a Transistor Digital. Felker..............................175–82
Computers and Switching Systems, Ferroelectric Materials as Storage Elements for Digital. Anderson..............................395–401
Computing Circuits and Devices for Industrial Process Functions. Hornfeck..................183–93
(Computing Devices) A Compact Inexpensive A-C Network Analyzer. Kimbark, Starr, Van Ness..............................122–7; disc. 127
(Computing Devices) A New Principle is Employed for 60-Cycle A-C Network Analyzers. Phillips..............................18–24; disc. 24
(Computing Devices) Design Improvements and Operating Experience with 10-Kc Network Analyzers. Ryder, Boast..................437–42; disc. 442
(Computing Devices) DINA, a Digital Analyzer for Laplace, Poisson, Diffusion, and Wave Equations. Leondes, Rubinoff.............303–08; disc. 308
(Computing Devices) Lighting a Network Analyzer Area. Dzwonczyk..............140–6; disc. 146
Constant Density of the Dielectric Breakdown of Paper with Air Resistance, The Variation at. Cloke, Khandelwal..................309–12; disc. 312
Constant Time Delay Lines, Design of Low-Frequency. Wallis..............................135–40
Control, Automatic Regulation of Metallic Rectifiers by Magnetic. Smith..................111–14
Control, Radio Interference. Maylott..........401–13
Core Materials on Magnetic Amplifier Circuits, The Effect of. Johnson..............26–30; disc. 30
Current in Multistage Magnetic Amplifiers, Compensating for the Quiescent. FitzGerald..........7–12
Current Status of Dynamic Stability Theory, The. Bothwell..............................223–8

D

Delay Lines, Design of Low-Frequency Constant Time. Wallis..............................135–40
Density of the Dielectric Breakdown of Paper with Air Resistance, The Variation at Constant. Cloke, Khandelwal..................309–12; disc. 312
Derivation of Vector Potential from Tables for Scalar Potential, The. Smith..................169–74
Design Improvements and Operating Experience with 10-Kc Network Analyzers. Ryder, Boast..............................437–42; disc. 442
Design of Eccles-Jordan Flip-Flops, Notes on the. Rubinoff..............................215–20
Design of Low-Frequency Constant Time Delay Lines. Wallis..............................135–40
Design of Unequal-Q Double-Tuned Transformers. Deutsch..............................314–20
Development of Telephony in the United States, The. Clark..............................348–64
Devices for Industrial Process Functions, Computing Circuits and. Hornfeck..................183–93
Diagrams for a Transistor Digital Computer, Typical Block. Felker..............................175–82
Dielectric Breakdown of Paper with Air Resistance, The Variation at Constant Density of the. Cloke, Khandelwal..................309–12; disc. 312
Dielectric, Electrical and Physical Properties of IN-420: A New Chlorinated Liquid. Warner.......330–5
Dielectric Heating at 915 Megacycles, Tubes for. Nelson..............................72–80
Dielectric Strength Characteristics of Liquid-Impregnated Pressboard, Impulse. Dakin, Works..............................321–5; disc. 325
Digital Analyzer for Laplace, Poisson, Diffusion, and Wave Equations, DINA, A. Leondes, Rubinoff..............................303–08; disc. 308
Digital Computer, Typical Block Diagrams for a Transistor. Felker..............................175–82
Digital Computers and Switching Systems, Ferroelectric Materials as Storage Elements for. Anderson..............................395–401
Dimensional Fields, A New Property of Two-. Moore..............................343–6; disc. 346
DINA, A Digital Analyzer for Laplace, Poisson, Diffusion, and Wave Equations. Leondes, Rubinoff..............................303–08; disc. 308
D-C Load, Saturable Reactors with Inductive. Storm: Part I. Steady-State Operation..........335–43
D-C Machine, Electromechanical Analogies of a Separately Excited. Vowels, Forte....32–41; disc. 41
Direct Method, Wave Filter Characteristics by a. Taylor, Watts..............................96–100
Direct-Scanning Facsimile System, A High-Speed. Deibert, Turner, Snider..................115–21
Direct-Writing Oscillograph, A High-Speed. Blumenthal, Eckels..............................50–6
Double-Tuned Transformers, Design of Unequal-Q. Deutsch..............................314–20
Dynamic Hysteresis Loop Measuring Equipment. Lord..............................269–72
Dynamic Stability Theory, The Current Status of. Bothwell..............................223–8

E

Eccles-Jordan Flip-Flops, Notes on the Design of. Rubinoff..........215–20
Effect of Core Materials on Magnetic Amplifier Circuits, The. Johnson..........26–30; disc. 30
Effective Feedback Ratio of Magnetic Amplifiers, The. Finzi, Pittman, Durand..........157–64
Electric Wave Filters, Semifinite Terminated. Chang..........149–56
Electrical and Physical Properties of IN-420: A New Chlorinated Liquid Dielectric. Warner....330–5
Electrical Engineering Problems by Southwell's Relaxation Method, Solution of. Grad..........205–14
Electromechanical Analogies of a Separately Excited D-C Machine. Vowels, Forte.....32–41; disc. 41
Electronic Recorder with Range and Precision Adequate for the Platinum Resistance Thermometer. Williams..........289–94; disc. 294
(Electronics) A Pulse Method for Supplying High-Voltage Power for Electrostatic Precipitation. White..........326–9; disc. 329
(Electronics) A Turns Index for Pulse Transformer Design. Lord..........165–8
(Electronics) Characteristic Shifts in Oxide Cathode Tubes. Bartley, White..........43–9
(Electronics) Tubes for Dielectric Heating at 915 Megacycles. Nelson..........72–80
Electrostatic Precipitation, A Pulse Method for Supplying High-Voltage Power for. White..........326–9; disc. 329
Element, The Single-Core Magnetic Amplifier as a Computer. Ramey..........442–6
Enterprise Upon Telegraphy's First Century, The Impress of. Coggeshall.........375–81; disc. 381
Equations, DINA, a Digital Analyzer for Laplace, Poisson, Diffusion and Wave. Leondes, Rubinoff..........303–08; disc. 308
Equipment, Dynamic Hysteresis Loop Measuring. Lord..........269–72
Excited D-C Machine, Electromechanical Analogies of a Separately. Vowels, Forte......32–41; disc. 41

F

Facsimile System, A High-Speed Direct-Scanning. Deibert, Turner, Snider..........115–21
Feedback Ratio of Magnetic Amplifiers, The Effective. Finzi, Pittman, Durand..........157–64
Ferroelectric Materials as Storage Elements for Digital Computers and Switching Systems. Anderson..........395–401
Fields, A New Property of Two-Dimensional. Moore..........343–6; disc. 346
Fifty Years of International Radio Communication, The First. Pratt..........371–5
Filter Characteristics by a Direct Method, Wave. Taylor, Watts..........96–100
Filters, Semifinite Terminated Electric Wave. Chang..........149–56
Flip-Flops, Notes on the Design of Eccles-Jordan. Rubinoff..........215–20
Fluid Mapper Patterns, Mapping Techniques Applied to. Moore..........1–5; disc. 5
Force, Calculation of the Magnetizing. Halacsy..90–5
Frequency Constant Time Delay Lines, Design of Low. Wallis..........135–40
Function of 2-Terminal Pair Networks, Synthesis of the Transfer. Kahal..........129–34
Functions, Computing Circuits and Devices for Industrial Process. Hornfeck..........183–93
Fundamental Accuracy of Single-Phase 3-Wire Metering, The. Wentz, Petzinger...295–302; disc. 302
Fundamental Plans for Toll Telephone Plant. Pilliod..........248–56

G

General-Purpose Relay for Telephone Switching Systems, A New. Keller..........413–28

H

Heating at 915 Megacycles, Tubes for Dielectric. Nelson..........72–80
High-Sensitivity Ratio Instrument for Industrial Resistance Thermometers, A. Clark..........57–60
High-Speed Direct-Scanning Facsimile System, A. Deibert, Turner, Snider..........115–21
High-Speed Direct-Writing Oscillograph, A. Blumenthal, Eckels..........50–6
High-Voltage Power for Electrostatic Precipitation, A Pulse Method for Supplying. White..........326–9; disc. 329
Historic Highlights in Developing the Radio Broadcasting and Television Arts. Hanson......364–71
Hysteresis Loop Measuring Equipment, Dynamic. Lord..........269–72
Hysteresis Loops, Ultrathin Tapes of Magnetic Alloys with Rectangular. Littmann..........220–3

I

Impedance of Rectangular Coaxial Transmission Lines, Characteristic. Omar, Miller..........81–9
Impregnated Pressboard, Impulse Dielectric Strength Characteristics of Liquid-. Dakin, Works..........321–5; disc. 325
Impress of Enterprise Upon Telegraphy's First Century, The. Coggeshall..........375–81; disc. 381
Improved Magnetic Servo Amplifier, An. Lufcy, Schmid, Barnhart..........281–9
Improvements and Operating Experience with 10-Kc Network Analyzers, Design. Ryder, Boast..........437–42; disc. 442
Impulse Dielectric Strength Characteristics of Liquid-Impregnated Pressboard. Dakin, Works..........321–5; disc. 325
IN-420: A New Chlorinated Liquid Dielectric, Electrical and Physical Properties of. Warner...330–5
Index for Pulse Transformer Design, A Turns. Lord..........165–8
Inductive D-C Load, Saturable Reactors with. Storm: Part I. Steady-State Operation..........335–43
Inductive Loading, Series-Connected Magnetic Amplifier with. Wilson..........101–10
Industrial Process Functions, Computing Circuits and Devices for. Hornfeck..........183–93
Industrial Resistance Thermometers, A High-Sensitivity Ratio Instrument for. Clark..........57–60
Inexpensive A-C Network Analyzer, A Compact. Kimbark, Starr, Van Ness.......122–7; disc. 127
Initial Capacitor Charge, Subharmonics in a Series Nonlinear Circuit as Influenced by. McKune, Brust..........200–05
Instrument for Industrial Resistance Thermometers, A High-Sensitivity Ratio. Clark..........57–60
(Instruments) A High-Speed Direct-Writing Oscillograph. Blumenthal, Eckels..........50–6
(Instruments) A Long-Distance Multipoint Telemetering System Using Teletype Transmission. Hornfeck, Markow..........193–9
(Instruments) Computing Circuits and Devices for Industrial Process Functions. Hornfeck...183–93
(Instruments) Electronic Recorder with Range and Precision Adequate for the Platinum Resistance Thermometer. Williams.......289–94; disc. 294
Instruments, Overload Protection of A-C. Pritchett, Trau..........13–17
Interference Control, Radio. Maylott..........401–13
International Radio Communication, The First 50 Years of. Pratt..........371–5

J

Jordan Flip-Flops, Notes on the Design of Eccles-Rubinoff..........215–20

L

Laplace, Poisson, Diffusion, and Wave Equations, DINA, A Digital Analyzer for. Leondes, Rubinoff..........303–08; disc. 308
Lines, Design of Low-Frequency Constant Time Delay. Wallis..........135–40
Liquid Dielectric, Electrical and Physical Properties of IN-420: A New Chlorinated. Warner.....330–5
Liquid-Impregnated Pressboard, Impulse Dielectric Strength Characteristics of. Dakin, Works..........321–5; disc. 325
Load, Saturable Reactors with Inductive D-C. Storm: Part I. Steady-State Operation..........335–43
Loading, Series-Connected Magnetic Amplifier with Inductive. Wilson..........101–10
Long-Distance Multipoint Telemetering System Using Teletype Transmission, A. Hornfeck, Markow..........193–9
Loop Measuring Equipment, Dynamic Hysteresis. Lord..........269–72
Loops, Ultrathin Tapes of Magnetic Alloys with Rectangular Hysteresis. Littmann..........220–3
Low-Frequency Constant Time Delay Lines, Design of. Wallis..........135–40

M

Machine, Electromechanical Analogies of a Separately Excited D-C. Vowels, Forte......32–41; disc. 41
Magnetic Alloys with Rectangular Hysteresis Loops, Ultrathin Tapes of. Littmann..........220–3
Magnetic Amplifier as a Computer Element, The Single-Core. Ramey..........442–6
Magnetic Amplifier Circuits, The Effect of Core Materials on. Johnson..........26–30; disc. 30
Magnetic Amplifier with Inductive Loading, Series-Connected. Wilson..........101–10
Magnetic Amplifiers, Compensating for the Quiescent Current in Multistage. FitzGerald..........8–12
(Magnetic Amplifiers) Dynamic Hysteresis Loop Measuring Equipment. Lord..........269–72
Magnetic Amplifiers of the Self-Balancing Potentiometer Type. Geyger..........383–95
(Magnetic Amplifiers) Saturable Reactors with Inductive D-C Load. Storm: Part. I. Steady-State Operation..........335–43
Magnetic Amplifiers—Survey, Types of. Miles..229–38
Magnetic Amplifiers, The Effective Feedback Ratio of. Finzi, Pittman, Durand..........157–64
Magnetic Amplifiers, The Figure of Merit of. Carleton, Horton..........239–45
Magnetic Control, Automatic Regulation of Metallic Rectifiers by. Smith..........111–14
Magnetic Servo Amplifier, A New Type of. Geyger..........272–80
Magnetic Servo Amplifier, An Improved. Lufcy, Schmid, Barnhart..........281–9
Magnetizing Force, The Calculation of the. Halacsy..........90–5
Mapping Techniques Applied to Fluid Mapper Patterns. Moore..........1–5; disc. 5
Materials as Storage Elements for Digital Computers and Switching Systems, Ferroelectric. Anderson..........395–401
Materials on Magnetic Amplifier Circuits, The Effect of Core. Johnson..........26–30; disc. 30
Measuring Equipment, Dynamic Hysteresis Loop. Lord..........269–72
Megacycles, Tubes for Dielectric Heating at 915. Nelson..........72–80
Merit of Magnetic Amplifiers, The Figure of. Carleton, Horton..........239–45
Metallic Rectifiers by Magnetic Control, Automatic Regulation of. Smith..........111–14
Metering, The Fundamental Accuracy of Single-Phase 3-Wire. Wentz, Petzinger.....295–302; disc. 302
Meters, Methods and Apparatus Employed in the Testing of Single-Phase A-C Watthour. Holtz..........61–9; disc. 69
Method for Supplying High-Voltage Power for Electrostatic Precipitation, A Pulse. White..........326–9; disc. 329
Method, Wave Filter Characteristics by a Direct. Taylor, Watts..........96–100
Methods and Apparatus Employed in the Testing of Single-Phase A-C Watthour Meters. Holtz..........61–9; disc. 69
Milestone in the Communication Arts, A. Bramhall..........37
Multipoint Telemetering System Using Teletype Transmission, A Long-Distance. Hornfeck, Markow..........193–9
Multistage Magnetic Amplifiers, Compensating for the Quiescent Current in. FitzGerald..........7–12

N

Nation-wide Numbering Plan. Nunn..........257–60
Nation-wide Telephone Service, Automatic Switching for. Clark, Osborne..........245–8
Network Analyzer, A Compact Inexpensive A-C. Kimbark, Starr, Van Ness........122–7; disc. 127
Network Analyzer Area, Lighting A. Dzwonczyk..........140–6; disc. 146
Network Analyzers, A New Principle Is Employed for 60-Cycle A-C. Phillips..........18–24; disc. 24
Network Analyzers, Design Improvements and Operating Experience with 10-Kc. Ryder, Boast..........437–42; disc. 442
Networks, Synthesis of the Transfer Function of 2-Terminal Pair. Kahal..........129–34
Nonlinear Circuit as Influenced by Initial Capacitor Charge, Subharmonics in a Series. McKune, Brust..........200–05
Notes on the Design of Eccles-Jordan Flip-Flops. Rubinoff..........215–20

O

Operating Experience with 10-Kc Network Analyzers, Design Improvements and. Ryder, Boast..........437–42; disc. 442
Oscillograph, A High-Speed Direct-Writing. Blumenthal, Eckels..........50–6
Overload Protection of A-C Instruments. Pritchett, Trau..........13–17
Oxide Cathode Tubes, Characteristic Shifts in. Bartley, White..........43–9

P

Pair Networks, Synthesis of the Transfer Function of 2-Terminal. Kahal..........129–34

Vol
Pt.
195

Paper with Air Resistance, The Variation at Constant Density of the Dielectric Breakdown of. Cloke, Khandelwal...309–12; disc. 312
Patterns, Mapping Techniques Applied to Fluid Mapper. Moore...1–5; disc. 5
Physical Properties of IN-420: A New Chlorinated Liquid Dielectric, Electrical and. Warner..330–5
Plan, Nationwide Numbering. Nunn...257–60
Plant, Fundamental Plans for Toll Telephone. Pilliod...248–56
Platinum Resistance Thermometer, Electronic Recorder with Range and Precision Adequate for the. Williams...289–94; disc. 294
Poisson, Diffusion, and Wave Equations, DINA, A Digital Analyzer for Laplace. Leondes, Rubinoff...303–08; disc. 308
Potential from Tables for Scalar Potential, The Derivation of Vector. Smith...169–74
Potentiometer Type, Magnetic Amplifiers of the Self-Balancing. Geyger...383–95
Power for Electrostatic Precipitation, A Pulse Method for Supplying High-Voltage. White...326–9; disc. 329
Precipitation, A Pulse Method for Supplying High-Voltage Power for Electrostatic. White...326–9; disc. 329
Precision Adequate for the Platinum Resistance Thermometer, Electronic Recorder with Range and. Williams...289–94; disc. 294
Pressboard, Impulse Dielectric Strength Characteristics of Liquid-Impregnated. Dakin, Works...321–5; disc. 325
Principle Is Employed for 60-Cycle A-C Network Analyzers, A New. Phillips...18–24; disc. 24
Problems by Southwell's Relaxation Method, Solution of Electrical Engineering. Grad...205–14
Process Functions, Computing Circuits and Devices for Industrial. Hornfeck...183–93
Properties of IN-420: A New Chlorinated Liquid Dielectric, Electrical and Physical. Warner..330–5
Property of Two-Dimensional Fields, A New. Moore...343–6; disc. 346
Protection of A-C Instruments, Overload. Pritchett, Trau...13–17
Pulse Method for Supplying High-Voltage Power for Electrostatic Precipitation, A. White...326–9; disc. 329
Pulse Transformer Design, A Turns Index for. Lord...165–8

Q

Quiescent Current in Multistage Magnetic Amplifiers, Compensating for the. FitzGerald...7–12

R

Radio Broadcasting and Television Arts, Historic Highlights in Developing the. Hanson...364–71
Radio Communication, The First 50 Years of International. Pratt...371–5
Radio Interference Control. Maylott...401–13
Range and Precision Adequate for the Platinum Resistance Thermometer, Electronic Recorder with. Williams...289–94; disc. 294
Ratio Instrument for Industrial Resistance Thermometers, A High-Sensitivity. Clark...57–60
Ratio of Magnetic Amplifiers, The Effective Feedback. Finzi, Pittman, Durand...157–64
Reactors with Inductive D-C Load, Saturable. Storm: Part I. Steady-State Operation...335–43
Recorder with Range and Precision Adequate for the Platinum Resistance Thermometer, Electronic. Williams...289–94; disc. 294
Rectangular Coaxial Transmission Lines, Characteristic Impedance of. Omar, Miller...81–9
Rectangular Hysteresis Loops, Ultrathin Tapes of Magnetic Alloys with. Littmann...220–3
Rectifiers by Magnetic Control, Automatic Regulation of Metallic. Smith...111–14
Regulation of Metallic Rectifiers by Magnetic Control, Automatic. Smith...111–14
Relaxation Method, Solution of Electrical Engineering Problems by Southwell's. Grad...205–14
Relay for Telephone Switching Systems, A New General-Purpose. Keller...413–28
Resistance Thermometers, A High Sensitivity Ratio Instrument for Industrial. Clark...57–60
Resistance Thermometer, Electronic Recorder with Range and Precision Adequate for the Platinum. Williams...289–94; disc. 294

S

Saturable Reactors with Inductive D-C Load. Storm: Part I. Steady-State Operation...335–43
Scalar Potential, The Derivation of Vector Potential from Tables for. Smith...169–74
Scanning Facsimile System, A High-Speed Direct-. Deibert, Turner, Snider...115–21
Self-Balancing Potentiometer Type, Magnetic Amplifiers of the. Geyger...383–95
Semifinite Terminated Electric Wave Filters. Chang...149–56
Separately Excited D-C Machine, Electromechanical Analogies of a. Vowels, Forte...32–41; disc. 41
Series-Connected Magnetic Amplifier with Inductive Loading. Wilson...101–10
Series Nonlinear Circuit as Influenced by Initial Capacitor Charge, Subharmonics in a. McKune, Brust...200–05
Servo Amplifier, A New Type of Magnetic. Geyger...272–80
Servo Amplifier, An Improved Magnetic. Lufcy, Schmid, Barnhart...281–9
Shifts in Oxide Cathode Tubes, Characteristic. Bartley, White...43–9
Single-Core Magnetic Amplifier as a Computer Element, The. Ramey...442–6
Single-Phase A-C Watthour Meters, Methods and Apparatus Employed in the Testing of. Holtz...61–9; disc. 69
Single-Phase 3-Wire Metering, The Fundamental Accuracy of. Wentz, Petzinger...295–302; disc. 302
Sixty-Cycle A-C Network Analyzers, A New Principle Is Employed for. Phillips...18–24; disc. 24
Solution of Electrical Engineering Problems by Southwell's Relaxation Method. Grad...205–14
Southwell's Relaxation Method, Solution of Electrical Engineering Problems by. Grad...205–14
Stability Theory, The Current Status of Dynamic. Bothwell...223–8
Status of Dynamic Stability Theory, The Current. Bothwell...223–8
Storage Elements for Digital Computers and Switching Systems, Ferroelectric Materials as. Anderson...395–401
Strength Characteristics of Liquid-Impregnated Pressboard, Impulse Dielectric. Dakin, Works...321–5; disc. 325
Subharmonics in a Series Nonlinear Circuit as Influenced by Initial Capacitor Charge. McKune, Brust...200–05
Survey—Types of Magnetic Amplifiers. Miles..229–38
Switching for Nationwide Telephone Service, Automatic. Clark, Osborne...245–8
Switching Systems, A New General-Purpose Relay for Telephone. Keller...413–28
Switching Systems, Automatic Toll. Shipley...261–9
Switching Systems, Ferroelectric Materials as Storage Elements for Digital Computers and. Anderson...395–401
Synthesis of the Transfer Function of 2-Terminal Pair Networks. Kahal...129–34
System, A High-Speed Direct-Scanning Facsimile. Deibert, Turner, Snider...115–21
Systems, Automatic Toll Switching. Shipley...261–9

T

Tables for Scalar Potential, The Derivation of Vector Potential from. Smith...169–74
Tapes of Magnetic Alloys with Rectangular Hysteresis Loops, Ultrathin. Littmann...220–3
Techniques Applied to Fluid Mapper Patterns, Mapping. Moore...1–5; disc. 5
(Telegraph) A High-Speed Direct-Scanning Facsimile System. Deibert, Turner, Snider...115–21
Telegraphy's First Century, The Impress of Enterprise Upon. Coggeshall...375–81; disc. 381
Telemetering System Using Teletype Transmission, A Long-Distance Multipoint. Hornfeck, Markow...193–9
Telephone Plant, Fundamental Plans for Toll. Pilliod...248–56
Telephone Service, Automatic Switching for Nationwide. Clark, Osborne...245–8
Telephone Switching Systems, A New General-Purpose Relay for. Keller...413–28
Telephone, Type-O Carrier. Coy, Van Tassel..428–37
Telephony in the United States, The Development of. Clark...348–64
Teletype Transmission, A Long-Distance Multipoint Telemetering System Using. Hornfeck, Markow...193–9
Television Arts, Historic Highlights in Developing the Radio Broadcasting and. Hanson...364–71
Terminal Pair Networks, Synthesis of the Transfer Function of 2-. Kahal...129–34
Terminated Electric Wave Filters, Semifinite. Chang...149–56
Testing of Single-Phase A-C Watthour Meters, Methods and Apparatus Employed in the. Holtz...61–9; disc. 69
Theory, The Current Status of Dynamic Stability. Bothwell...223–8
Thermometer, Electronic Recorder with Range and Precision Adequate for the Platinum Resistance. Williams...289–94; disc. 294
Thermometers, A High-Sensitivity Ratio Instrument for Industrial Resistance. Clark...57–60
Three-Wire Metering, The Fundamental Accuracy of Single-Phase. Wentz, Petzinger...295–302; disc. 302
Time Delay Lines, Design of Low-Frequency Constant. Wallis...135–40
Toll Switching Systems, Automatic. Shipley...261–9
Toll Telephone Plant, Fundamental Plans for. Pilliod...248–56
Transfer Function of 2-Terminal Pair Networks, Synthesis of the. Kahal...129–34
Transformer Design, A Turns Index for Pulse. Lord...165–8
Transformers, Design of Unequal-Q Double-Tuned. Deutsch...314–20
Transistor Digital Computer, Typical Block Diagrams for a. Felker...175–82
Transmission, A Long-Distance Multipoint Telemetering System Using Teletype. Hornfeck, Markow...193–9
Transmission Lines, Characteristic Impedance of Rectangular Coaxial. Omar, Miller...81–9
Tubes, Characteristic Shifts in Oxide Cathode. Bartley, White...43–9
Tubes for Dielectric Heating at 915 Megacycles. Nelson...72–80
Tuned Transformers, Design of Unequal-Q Double-. Deutsch...314–20
Turns Index for Pulse Transformer Design, A. Lord...165–8
Two-Dimensional Fields, A New Property of. Moore...343–6; disc. 346
Type-O Carrier Telephone. Coy, Van Tassel...428–37
Types of Magnetic Amplifiers—Survey. Miles...229–38
Typical Block Diagrams for a Transistor Digital Computer. Felker...175–82

U

Unequal-Q Double-Tuned Transformers, Design of. Deutsch...314–20
Ultrathin Tapes of Magnetic Alloys with Rectangular Hysteresis Loops. Littmann...220–3
United States, The Development of Telephony in the. Clark...348–64

V

Variation at Constant Density of the Dielectric Breakdown of Paper with Air Resistance, The. Cloke, Khandelwal...309–12; disc. 312
Vector Potential from Tables for Scalar Potential, The Derivation of. Smith...169–74

W

Watthour Meters, Methods and Apparatus Employed in the Testing of Single-Phase A-C. Holtz...61–9; disc. 69
Wave Equations, DINA, a Digital Analyzer for Laplace, Poisson, Diffusion, and. Leondes, Rubinoff...303–08; disc. 308
Wave Filter Characteristics by a Direct Method. Taylor, Watts...96–100
Wave Filters, Semifinite Terminated Electric. Chang...149–56

2. Author Index

Anderson, J. R. Ferroelectric Materials as Storage Elements for Digital Computers and Switching Systems. (52-309)...395–401

B

Barnhart, P. W.; C. W. Lufcy, A. E. Schmid. An Improved Magnetic Servo Amplifier. (52-235)...281–9
Bartley, William P.; J. E. White. Characteristic Shifts in Oxide Cathode Tubes. (52-53)...43–9
Blumenthal, I. S.; A. R. Eckels. A High-Speed Direct-Writing Oscillograph. (52-54)...50–6
Boast, W. B.; J. D. Ryder. Design Improvements and Operating Experience with 10-Kc Network Analyzers. (52-310)...437–42; disc. 442
Bothwell, Frank E. The Current Status of Dynamic Stability Theory. (52-214)...223–8
Bramhall, F. B. A Milestone in the Communication Arts...347

Brust, Melvin F.; W. J. McKune. Subharmonics in a Series Nonlinear Circuit as Influenced by Initial Capacitor Charge. (52-207)............200–05
Busch, F. H. Disc..............................70, 302

C

Carleton, J. T.; W. F. Horton. The Figure of Merit of Magnetic Amplifiers. (52-234)........239–45
Chang, Sheldon S. L. Semifinite Terminated Electric Wave Filters. (52-123)................149–56
Clark, A. B. The Development of Telephony in the United States. (52-258)................348–64
Clark, A. B.; H. S. Osborne. Automatic Switching for Nation-wide Telephone Service. (52-201)........245–8
Clark, E. W. A High-Sensitivity Ratio Instrument for Industrial Resistance Thermometers. (52-55)........57–60
Cloke, Paul; K. K. Khandelwal. The Variation at Constant Density of the Dielectric Breakdown of Paper with Air Resistance. (52-210)........309–12; disc. 314
Coggeshall, I. S. The Impress of Enterprise Upon Telegraphy's First Century. (52-257).....375–81
Coy, J. A.; E. K. Van Tassel. Type-O Carrier Telephone. (52-307)................428–37
Cybulski, John. Disc..............................41

D

D'Andrade, H. E. Disc..............................148
Dakin, T. W.; C. N. Works. Impulse Dielectric Strength Characteristics of Liquid-Impregnated Pressboard. (52-228)...........321–5; disc. 325
Deibert, C. R.; F. T. Turner, R. H. Snider. A High-Speed Direct-Scanning Facsimile System. (52-103)........115–21
Del Mar, William A. Disc..............................312
Deutsch, Sid. Design of Unequal-Q Double-Tuned Transformers. (52-212)................314–20
Douglas, John F. H. Disc..............................6, 346
Durand, H. L.; L. A. Finzi, G. F. Pittman, Jr. The Effective Feedback Ratio of Magnetic Amplifiers. (52-124)..............................157–64
Dzwonczyk, V. L. Lighting a Network Analyzer Area. (52-119)....................140–6; disc. 148

E

Eckels, A. R.; I. S. Blumenthal. A High-Speed Direct Writing Oscillograph. (52-54)............50–6

F

Felker, J. H. Typical Block Diagrams for a Transistor Digital Computer. (52-162)............175–80
Field, William J. Disc..............................32
Finzi, L. A.; G. F. Pittman, Jr., H. L. Durand. The Effective Feedback Ratio of Magnetic Amplifiers. (52-124)..............................157–64
FitzGerald, Alan S. Compensating for the Quiescent Current in Multistage Magnetic Amplifiers. (52-5)........7–12
Forte, W. G.; R. E. Vowels. Electromechanical Analogies of a Separately Excited D-C Machine. (52-45)....................32–41; disc. 41
Franck, Kurt. Disc..............................148

G

Geyger, W. A. A New Type of Magnetic Servo Amplifier. (52-237)..............................272–80
Geyger, W. A. Magnetic Amplifiers of the Self-Balancing Potentiometer Type. (52-271).....383–95
Grad, E. M. (Miss). Solution of Electrical Engineering Problems by Southwell's Relaxation Method. (52-213)..............................205–14

H

Halacsy, A. A. The Calculation of the Magnetizing Force. (52-84)..............................90–5
Hanson, O. B. Historic Highlights in Developing the Radio Broadcasting and Television Arts. (52-261)........364–71
Hirst, J. M. Disc..............................127
Holtz, F. C. Methods and Apparatus Employed in the Testing of Single-Phase A-C Watthour Meters. (52-56)....................61–9; disc. 71
Hornfeck, A. J. Computing Circuits and Devices for Industrial Process Functions. (52-191)....183–93
Hornfeck, A. J.; G. R. Markow. A Long-Distance Multipoint Telemetering System Using Teletype Transmission. (52-197)................193–9
Horton, W. F.; J. T. Carleton. The Figure of Merit of Magnetic Amplifiers. (52-234)..........239–45
Howell, F. H. Disc..............................147

J

Johnson, Leo J. The Effect of Core Materials on Magnetic Amplifier Circuits. (52-26)..26–30; disc. 30

K

Kahal, Robert. Synthesis of the Transfer Function of 2-Terminal Pair Networks. (52-146)......129–34
Kahler, W. H. Disc..............................147
Keller, A. C. A New General-Purpose Relay for Telephone Switching Systems. (52-304)..........413–28
Khandelwal, Krishna K.; P. Cloke. The Variation at Constant Density of the Dielectric Breakdown of Paper with Air Resistance. (52-210)........309–12; disc. 314
Kimbark, Edward W.; J. H. Starr, J. E. Van Ness. A Compact Inexpensive A-C Network Analyzer. (52-113)....................122–7; disc. 128
Kraus, John D. Disc..............................5

L

Le Vesconte, Lester B. Disc..............................127
Lenehan, B. E. Disc..............................70
Leondes, Cornelius; M. Rubinoff. DINA, a Digital Analyzer for Laplace, Poisson, Diffusion and Wave Equations. (52-181)..........303–08; disc. 308
Lindsay, E. A. Disc..............................147
Littmann, Martin F. Ultrathin Tapes of Magnetic Alloys with Rectangular Hysteresis Loops. (52-208)........220–3
Lord, Harold W. A Turns Index for Pulse Transformer Design. (52-125)....................165–8
Lord, H. W. Dynamic Hysteresis Loop Measuring Equipment. (52-238)....................269–72
Lufcy, C. W.; A. E. Schmid, P. W. Barnhart. An Improved Magnetic Servo Amplifier. (52-235)........281–9

M

Markow, G. R.; A. J. Hornfeck. A Long-Distance Multipoint Telemetering System Using Teletype Transmission. (52-197)................193–9
Maylott, Carleton F. Radio Interference Control. (52-300)..............................401–13
McKune, William J.; M. F. Brust. Subharmonics in a Series Nonlinear Circuit as Influenced by Initial Capacitor Charge. (52-207)............200–05
Miles, James G. Types of Magnetic Amplifiers—Survey. (52-216)....................229–38
Miller, C. Frank; Y. A. Omar. Characteristic Impedance of Rectangular Coaxial Transmission Lines. (52-75)..............................81–9
Moore, A. D. A New Property of Two-Dimensional Fields. (52-240)....................343–6
Moore, A. D. Mapping Techniques Applied to Fluid Mapper Patterns. (52-3)..........1–5; disc. 6

N

Nelson, Richard B. Tubes for Dielectric Heating at 915 Megacycles. (52-66)....................72–80
Nunn, W. H. Nation-wide Numbering Plan. (52-203)........257–60

O

Omar, Younes A.; C. F. Miller. Characteristic Impedance of Rectangular Coaxial Transmission Lines. (52-75)..............................81–9
Osborne, H. S.; A. B. Clark. Automatic Switching for Nation-wide Telephone Service. (52-201).245–8

P

Petzinger, A. J.; E. C. Wentz. The Fundamental Accuracy of Single-Phase 3-Wire Metering. (52-163)....................295–302; disc. 302
Phillips, E. B. A New Principle is Employed for 60-Cycle A-C Network Analyzers. (52-19)........18–24; disc. 25
Pilliod, J. J. Fundamental Plans for Toll Telephone Plant. (52-202)....................248–56
Pittman, G. F., Jr.; L. A. Finzi, H. L. Durand. The Effective Feedback Ratio of Magnetic Amplifiers. (52-124)..............................157–64
Pratt, Haraden. The First 50 Years of International Radio Communication. (52-263)..........371–5
Pritchett, Wilson; E. Trau. Overload Protection of A-C Instruments. (52-7)....................13–17

R

Ramey, R. A. The Single-Core Magnetic Amplifier as a Computer Element. (52-293)..........442–6
Rubinoff, Morris. Notes on the Design of Eccles-Jordan Flip-Flops. (52-215)............215–20
Rubinoff, Morris; C. Leondes. DINA, a Digital Analyzer for Laplace, Poisson, Diffusion and Wave Equations. (52-181)..........303–08; disc. 308
Ryder, J. D.; W. B. Boast. Design Improvements and Operating Experience with 10-Kc Network Analyzers. (52-310)..............437–42; disc. 442

S

Schleicher, G. B. Disc..............................71
Schmid, A. E.; C. W. Lufcy, P. W. Barnhart. An Improved Magnetic Servo Amplifier. (52-235)........281–9
Shipley, F. F. Automatic Toll Switching Systems. (52-204)..............................261–9
Silsbee, F. B. Disc..............................69
Smith, D. H. Automatic Regulation of Metallic Rectifiers by Magnetic Control. (52-102)......111–14
Smith, J. J. The Derivation of Vector Potential from Tables for Scalar Potential. (52-129).....169–74
Snider, R. H.; C. R. Deibert, F. T. Turner. A High-Speed Direct-Scanning Facsimile System. (52-103)........115–21
Sommerman, G. M. L. Disc..............................325
Squires, R. B. Disc..............................24, 127, 442
Starr, James H.; E. W. Kimbark, J. E. Van Ness. A Compact Inexpensive A-C Network Analyzer. (52-113)....................122–7; disc. 128
Storm, H. F. Saturable Reactors with Inductive D-C Load: Part I. Steady-State Operation. (52-236)........335–43
Stull, Daniel R. Disc..............................295

T

Taylor, George J. Disc..............................146
Taylor, R. C.; Mrs. C. U. Watts. Wave Filter Characteristics by a Direct Method. (52-85)...96–100
Thomas, E. R. Disc..............................329
Trau, Elazar; W. Pritchett. Overload Protection of A-C Instruments. (52-7)....................13–17
Treves, Jacques. Disc..............................381
Turner, F. T.; C. R. Deibert, R. H. Snider. A High-Speed Direct-Scanning Facsimile System. (52-103)........115–21

V

Van Ness, James E.; E. W. Kimbark, J. H. Starr. A Compact Inexpensive A-C Network Analyzer. (52-113)....................122–7; disc. 128
Van Tassel, E. K.; J. A. Coy. Type-O Carrier Telephone. (52-307)....................428–37
Vanderleck, J. M. Disc..............................70
Vowels, R. E.; W. G. Forte. Electromechanical Analogies of a Separately Excited D-C Machine. (52-45)....................32–41; disc. 41

W

Waite, G. G. Disc..............................71
Wallis, Clifford M. Design of Low-Frequency Constant Time Delay Lines. (52-150)......135–40
Warner, A. J. Electrical and Physical Properties of IN-420: A New Chlorinated Liquid Dielectric. (52-227)..............................330–5
Watts, Mrs. C. U.; R. C. Taylor. Wave Filter Characteristics by a Direct Method. (52-85).......96–100
Wentz, E. C.; A. J. Petzinger. The Fundamental Accuracy of Single-Phase 3-Wire Metering. (52-163)....................295–302; disc. 302
Whinnery, J. R. Disc..............................308
White, H. J. A Pulse Method for Supplying High-Voltage Power for Electrostatic Precipitation. (52-230)....................326–9; disc. 330
White, John E.; W. P. Bartley. Characteristic Shifts in Oxide Cathode Tubes. (52-53)..........43–9
Williams, Albert J., Jr. Electronic Recorder with Range and Precision Adequate for the Platinum Resistance Thermometer. (52-157)...289–94; disc. 295
Wilson, Thomas G. Series-Connected Magnetic Amplifier with Inductive Loading. (52-94)....101–10
Witschonke, C. R. Disc..............................294
Works, C. N. Disc..............................313
Works, C. N.; T. W. Dakin. Impulse Dielectric Strength Characteristics of Liquid-Impregnated Pressboard. (52-228).........321–5; disc. 325

Mapping Techniques Applied to Fluid Mapper Patterns

A. D. MOORE
FELLOW AIEE

MANY field problems calling for graphical mapping have been crudely solved, due to lack of graphical skill or lack of training. Other problems have sometimes been badly solved by using mathematics on simplified but inaccurate substitutes for the real conformation. Many good solutions have been obtained with the electrolytic tank.

With the advent of fluid mappers, many fluid-flow simulations of magnetic and other analogous fields become possible. Papers on fluid mapper techniques are cited in the references.

A fluid mapper gives the first step in solving a field problem by furnishing a flow pattern with a large number of visible lines. The completion of a map based on some of the lines, and the analysis of the map for solution purposes, involve nothing more than simple graphics and arithmetic. This paper tells how to take these steps. Also, a number of the techniques apply equally well to graphical or other types of mapping.

The method of circling-in good curvilinear squares is presented, along with studies showing accuracies achieved. The more important properties of curvilinear square volumes are described. Methods for mapping and evaluating tubes are set forth, and the use of the map for further analysis is discussed.

Until late in the paper, the content is concerned only with 2-dimensional situations. Then follows a discussion of the 3-dimensional field with axial symmetry, and the quasi-3-dimensional field.

The Field Map

At the least, a field map will give the observer a general idea of the field conformation and a rough judgment of values. At best, it accurately shows the conformation and can be used to solve all parts of the field problem. That is, the map can be analyzed for numerical results.

Most published field maps have come by way of mathematics, by graphical mapping, or from the electrolytic tank. Thus, nearly all 2-dimensional maps have shown a regular array of curvilinear squares. Such maps inevitably have influenced our thinking, and lead us to expect to construct such a map when starting from a fluid mapper pattern. This expectation is not in order. In using fluid mapper patterns, one learns to deal with arbitrary tubes of generally unequal values.

The Curvilinear Square

The important entity in field mapping is the curvilinear square. Consider a curvilinear rectangle, which is any area bounded by two flow lines and two equipotential lines. Subdivide the area by means of the two midlines: one divides the potential in halves, the other divides the flow in halves. The result is four curvilinear rectangles. Likewise, subdivide each of the four, making 16. Continue to the limit, at which all would approach true rectangles, all having the same length-width ratio. This is regular subdivision.

If such an area is regularly subdivided to the limit and the infinitesimal areas approach true squares, then the area was a curvilinear square.

The Curvilinear Square Volume

Think of current in the tube shown in Figure 1(B). At the left end is curvilinear square ABCD. The map has only two dimensions, and we must visualize the third dimension, normal to the paper. ~~Thus the square~~ ABCD really represents a curvilinear square volume. Equipotential lines AB and CD represent equipotential surfaces. Current flows in the volume from one surface to the other. Current lines AD and BC really represent current sheets.

In the geometric sense, a curvilinear square volume is generated by translating a curvilinear square normal to its plane, through the distance of the third dimension. The rather lengthy term curvilinear square volume might be replaced by the word cell for the sake of brevity. Professor John Kraus of Ohio

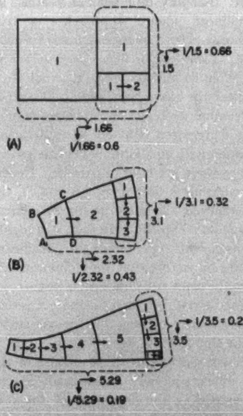


Figure 1. Illustrating the mapping of a tube with curvilinear squares plus a remainder, and treatment of the remainder; examples showing how to find the relative ohms of the tube

---

Paper 52-3, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21-25, 1952. Manuscript submitted February 15, 1951; made available for printing November 13, 1951.

A. D. MOORE is with the University of Michigan, Ann Arbor, Mich.

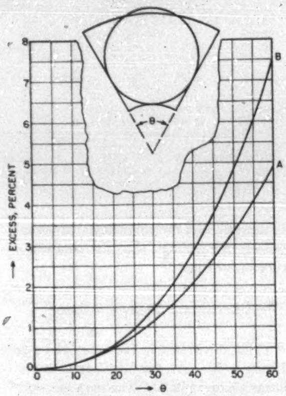


Figure 2. Error, when sectorial squares are circled in, expressed as an excess

Curve A: excess, as per cent of a curvilinear square

Curve B: excess, as per cent of circle diameter

State University has started this practice, and the writer thinks well of it.

Properties of Curvilinear Square Volumes

Such volumes have three very important properties. The first may be called the property of inversion. Think of regularly subdividing volume ABCD to the limit. For any infinitesimal volume, its curvilinear square approaches a true square, and its flow approaches rectilinear flow. Therefore, if flow is set up across the volume either way, it would encounter the same resistance. All of the little volumes have the same resistance. The aggregation is a series-parallel combination with equal numbers in series in a row as there are rows in parallel. Hence, for curvilinear flow from AB to CD, the resistance is the same as for curvilinear flow from AD to BC. In fact, it is highly advantageous to define a curvilinear square by saying that the volume would have the same resistance each way.

The second property is that of resistance equality among the various curvilinear square volumes of a map. This is true, irrespective of shape or size of the curvilinear squares. This is easily proved by extending the argument of the preceding paragraph.

The third property, by like argument relates the resistance of a curvilinear square volume to that of a unit cube. Rectilinear flow through a unit cube encounters the same resistance as does flow through a curvilinear square volume of unit axial dimension, the third dimension. Since, for a known medium, we will know the resistance, or permeance, et cetera, of the unit cube, we can then evaluate the absolute ohms of a curvilinear square volume.

Table I. The Excess, When Sectorial Squares Are Circled In

θ, Degrees	Excess, as % of a Curvilinear Square	Excess, as % of Circle Diameter
10	0.15	0.16
20	0.52	0.60
30	1.16	1.48
40	2.10	2.84
50	3.33	4.82
60	4.91	7.52

Absolute Resistance

Let a tube have ten curvilinear square volumes in series. Let the medium have a resistivity of 6×10^{-6} ohm for a unit cube. If the axial or third dimension is one unit, the resistance of a curvilinear square volume is 6×10^{-6}. The tube resistance, ten times as much, is 6×10^{-5}. Or, if the axial dimension is three units, tube resistance is 6×10^{-5} divided by three, or 2×10^{-5}. Thus absolute resistances are found. Before a map can be numerically analyzed, there must be absolute evaluations; but this step often can be put off until near the end, since earlier work is often best done in terms of relative ohms.

Relative Ohms

In the concept of the relative ohm, we simply assign one relative resistance unit to each curvilinear square volume, irrespective of the medium and the axial dimension. Relative resistance units may as well be called relative ohms, for short.

Tube Evaluation in Terms of Relative Ohms

In Figure 1, all three tubes have been mapped and evaluated. Beginning at one end, a series of full squares is laid out, continuing until a remainder is left. Beginning at one end of the remainder, another series of squares is laid out until a second remainder is left. It can be treated as before; in theory, the process could carry on forever. In practice, the second remainder can often be estimated well enough, without mapping it. Many times, the first remainder can be so estimated. The result is a series-parallel combination of unit values. To evaluate, we begin with the last remainder and operate by a series of inversions. For all three tubes in Figure 1, the steps and the values are shown. Only one case need be taken up here.

In Figure 1(B) the second remainder is unmapped and is estimated as a 1 by 10 curvilinear rectangle. That is, it is estimated that ten little curvilinear squares could be mapped into it. Invert, and get 0.1 relative ohm. Place in series with the three squares above it and get, for downward flow, 3.1 relative ohms. Invert, and for rightward flow the combination so far gives 1/3.1 or 0.32 relative ohm. This, in series with the two large squares to the left, gives 2.32 relative ohms for rightward flow in the tube. Downward flow across the tube would encounter 1/2.32, or 0.43 relative ohm.

Constructing Curvilinear Squares

The fastest way to construct curvilinear squares is simply to sketch them in. A trained worker can do remarkably well. However, illusion is present, and one's judgment is sometimes badly fooled. While teaching graphical mapping for over three decades, the writer long since began trying to devise various graphical and mechanical aids for constructing curvilinear squares. The only one of these having much promise was a circle method, but it remained on ice until a few years ago for lack of a suitable tool. When the Lietz Circle Guide (The A. Lietz Company) appeared, the method became feasible. The writer then developed the circling-in method, along with the accuracy studies pertaining to it as included herein. It has been thoroughly tried out by hundreds of students. Even the poor free-hand mapper can do good and fast work with this circling-in method.

It is an approximate method. To hunt for an accurate method is futile. Due to the infinite variety of possible shapes of curvilinear squares, there is not and never can be a simple, fast, and at the same time accurate method for constructing squares or for checking them after they are constructed.

The Circling-in Method

To map a tube, we typically start at one end and fill in with full curvilinear squares, until a part-square or remainder is left at the other end. The typical task is: given three boundaries, to locate and draw the fourth.

To circle-in a square, select a circle of the most nearly correct size in the Lietz tool, move the circle into place, and lightly draw a circle tangent to the three boundaries. The fourth side is then faired in orthogonally to follow local trends and to be tangent to the circle. The number of circular guide holes is limited. If too small a circle must be used, the work is done so as to make the circle equidistant from the four tangency points; likewise, for a too-large circle. Since the method is approximate, the question must be raised, how good is it?

Sectorial Squares: Accuracy of the Circling-in Method

Correct sectorial squares, bounded by straight radials and circular arcs, are easily obtained mathematically. One such for a 60-degree angle is shown in Figure 2. We see that a circle fitted to the inner arc and to the radials extends beyond the outer arc: there is an excess. Certain excess values have been accurately worked out and presented in two ways in Table I. For example, for the above angle the excess in per cent of a curvilinear square is 4.91 per cent. This means that if an approximate square is circled-in for this angle, it is really a curvilinear rectangle. For radial flow it would have 1.0491 relative ohms, rather than unit value.

For the same angle, the table also shows that the excess in per cent of circle diameter is 7.52 per cent. Thus, a correct square could be drawn by drawing the circle, then coming radially inward from its outermost point by a distance equal to 7.52 per cent of the diameter to locate the outer arc.

The tabular data are given in curve form, Figure 2. Curve A shows that at 38 degrees an uncorrected square, as circled-in, has an error of only 2 per cent; at 27.5 degrees, only 1 per cent.

Curvilinear Squares Having Midline Symmetry

In teaching, the writer uses the term midline to mean a flow line dividing the flow in halves, or an equipotential line dividing the potential in halves. When a curvilinear square is regularly subdivided, it is the midlines that are used. A square having symmetry about one of its midlines may be said to have midline symmetry. Any number of shapes is possible, only two being shown in Figure 3. Having symmetry about vertical midlines, these were constructed by the writer by graphical synthesis to have quite good accuracy. In the upper shape it is seen that a circle fitted in falls short of the outer arc: there is a deficit. The lower shape shows a very considerable excess.

These cases point up a warning: large error can enter in if highly distorted squares are circled-in. The usual remedy would be first to subdivide, circling-in the four parts; sometimes, further work on one or two of the parts may be needed.

Fortunately, there is a very broad exception to this need for subdividing distorted squares. It is considered next.

Curvilinear Squares Having Diagonal Symmetry

In a true square, the diagonals are straight. In a curvilinear square the diagonals are generally curved. But if one of them is straight, and the square is symmetrical about it, we may say that the square has diagonal symmetry. It is obvious that a circle can be fitted perfectly into any such square, with tangency at all four points. It is fortunately true of a great many fields that many of their curvilinear squares very nearly achieve diagonal symmetry. The circling-in method for constructing such squares has high accuracy.

Fluid Mapper Pattern

A pattern is shown in Figure 4. This 8 by 12 inch slab has a well 2.36 inches in diameter. This is a picture of a record slab rather than of the actual visible flow lines.

The slab was coated with Spred-Satin,[4] dye lines formed from potassium permanganate crystals reacted chemically with the paint and formed the permanent flow record. A large number of good flow lines were thus made available. Only 33 lines were traced from the record slab to make the basis for the Figure 5 map.

Selection of Lines to Be Traced

Up to now, those familiar with field maps usually expect a map to show a regular array of squares, with both sets of lines showing as continuous lines. Many are frankly puzzled when faced with the numerous and random lines presented by a fluid mapper pattern. How do we know which lines to select to make the map come out right? We do not need to know.

Corner lines numbered 0, 6, 18 and 22 are drawn, to avoid dealing with what are often called "5-sided squares." Remaining lines are selected to form tubes easily subjected to mapping, and so that nearly all squares can be circled-in with acceptable accuracy.

Therefore, line selection is an arbitrary matter, leading to a collection of unequal tubes. Only by accident will the relative ohms of two of the tubes be the same.

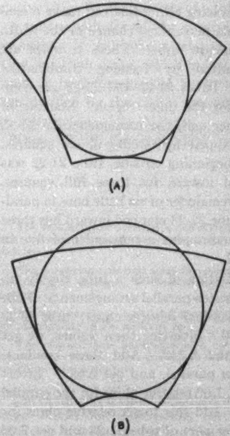


Figure 3 (below, left). Two shapes of curvilinear squares having midline symmetry. If (A) is circled in, it has a small deficit error. If (B) is circled in, it has a large excess error

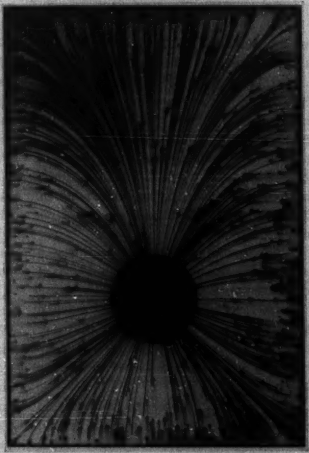


Figure 4 (below, right). A record slab. The permanent record of a flow pattern made in a fluid mapper. A tracing of some of the lines initiated the map of Figure 5

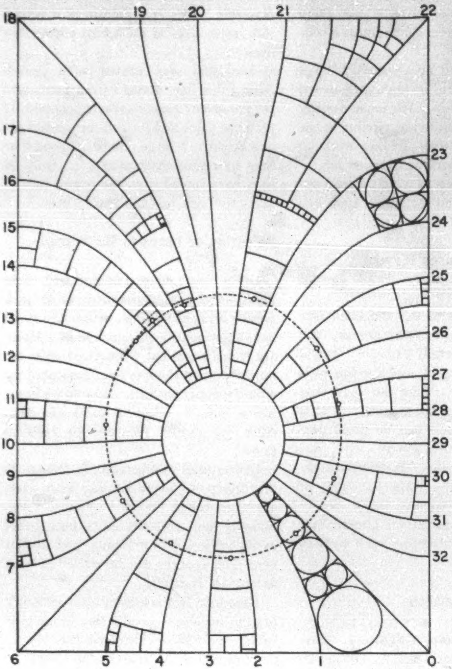


Figure 5. A field map, based on the record slab pattern, and drawn to show the application of a number of mapping techniques, also showing a derived 50 per cent equipotential line

Mapping the Tubes

By leaving some tubes unmapped in Figure 5, the mapped tubes stand out more clearly, and procedures are more apparent.

Tube 0-1 was mapped from both ends, with a remainder left somewhere where the two processes meet in the tube. This is permissible and sometimes desirable. Starting inside, four squares were circled-in. Starting outside, the large square is highly distorted. Even so, it has almost perfect diagonal symmetry and can be circled-in as a whole. The remainder happens to amount to two little squares. By inspection, this tube has 5.5 relative ohms.

A great many of the squares in this map will be seen to have a very high degree of diagonal symmetry.

With two exceptions, all other mapped tubes were circled-in starting at the inner ends. By inspection, tube 13-14 has six relative ohms; tube 14-15 has 10.5; tube 27-28 has 6.2.

In a field region where tubes taper from very broad to very narrow, a special problem arises. If the two lines selected to form a tube are spaced for proper mapping at the wide end, their spacing is too narrow at the other end for two reasons. First, it takes time to fill in numerous little squares. Second, a very narrow part of a tube may be inaccurate: a small error in tracing either line can make a considerable percentage change in the width of a narrow tube. These troubles are surmounted by forming combination tubes. Tubes 21-22 and 22-23 are combined for the inner region. Starting at the inner end, the combined tube 21-23 was mapped by circling-in six squares. Next, beginning outside, tube 21-22 was mapped inward for three full squares, plus a remainder of six little ones in parallel. Tube 22-23 mapped inward has three full squares plus a remainder of five in parallel.

Evaluation of such a tube simply involves series-parallel arrangements. Tube 21-22 has 3.17 relative ohms; tube 22-23 has 3.20. Inverting these values, we get 0.315 and 0.312. Add these conductances in parallel, and get 0.627. Invert and get 1.60 relative ohms for the parallel paths. Add this to six relative ohms for the inner part of tube 21-23 and get 7.60 for the entire combined tube. Tube 16-17 and tube 17-18 have been similarly treated.

Partial Subdivision

A curvilinear square with a corner terminating at a singular point or at a boundary corner may need special treatment. The large square at the outer end of tube 21-22 offers such a case. Think of flow density arriving at the 21-22 boundary. Going from point 21, it decreases, approaching zero at point 22. Therefore, to find numerical values for such a thing as flux density in a magnetic field along such a boundary, or within such a square, an over-all average value would not do: the square needs subdivision. But to carry out several steps of complete regular subdivision would involve much needless work.

Instead, what the writer calls partial subdivision, or subdivision toward a corner, is carried out. This square has been regularly subdivided once, into four squares. Then, taking only the square in the corner, it is subdivided once. Then, of the four new squares constructed, subdivide only the corner square, and so on. In theory, the process goes to an infinite number of successive partial subdivisions.

Continue to think of magnetic flux in this tube. Of all this flux, half is carried by the first square to the right of point 21; one-fourth, by the next little square; one-eighth by the next, and so on. Knowing this, and by measuring widths, highly accurate localized average densities can be found.

Deriving an Equipotential Line

The map can be made to yield any desired equipotential line. To illustrate, the 50 per cent equipotential is found in Figure 5, each mapped tube having been made to help locate it. Tube 13-14 happened to map without a remainder, and it has six relative ohms. Going out through three squares takes in half the relative ohms and therefore half the potential difference. The little circle located on the arc shows where the desired line should locate, according to this tube. Next take tube 14-15, with 10.5 relative ohms. Half of this is 5.25: we must go out through 5.25 curvilinear squares. This calls for interpolating within a square. The full process could be to subdivide the sixth square twice into 16 little squares. Usually it is sufficient to sketch in the midlines, shown dotted, and then interpolate by estimate.

Each mapped tube is thus allowed to

help locate the 50 per cent equipotential. When the line, shown dashed, is faired in, it is seen here to have orthogonality and to go through or very close to all points. Thus, the map's accuracy has been checked by internal evidence.

Flow Analysis

In addition to finding any desired potential, map analysis can be made to find the entire flow or any part of it. Suppose Figure 5 were completed by mapping all tubes. After counting a tube's relative ohms, the value is inverted to get relative conductance. All conductances are added to get total relative conductance; and this, inverted, gives total relative ohms. By bringing in the absolute resistance of the unit cube, and the axial dimension, the total absolute ohms can be found. Voltage applied between inner and outer boundaries, divided by the ohms, gives the total current.

Suppose it is desired to divide the flow into eight equal parts. Divide the total relative conductance by eight to find relative conductance for each part. Starting somewhere, go across tubes, adding their relative conductances. A place will be reached where, to add in the next tube would give too much conductance. It is easy to compute what fraction of this tube to take in to make up the desired amount of conductance.

Three-Dimensional Fields Having Axial Symmetry

Fluid mappers can solve some cases having axial symmetry. The flow space is built to vary directly or inversely with the cube root of the radius, depending on whether it is for the pro-axis or cross-axis type. But once the mapper is operated and the pattern obtained, the flow space variation must be put out of mind: in the analysis, think only of the pattern lines as belonging to the physical case being solved.

As before, lines from the pattern are arbitrarily selected for tracing, to form tubes that can readily be mapped. Again, a tube is mapped full of squares by the circling-in method. Now, in the 2-dimensional work, each square could have one relative ohm assigned, since all volumes had the same third dimension; but in axial symmetry work, the third dimensions differ, they being proportional to the squares' radii.

Also, since each square actually has any number of radii, and since the radii for different points within a square near to or touching the axis vary enormously, the selection of only one radius to use for a given square becomes a very important matter. The writer's practice will be outlined.

First, subdivide a square once, just as if it were part of a 2-dimensional field, dividing the square into four curvilinear squares. Note that these are not true midlines for the real field and must not be mistaken for such. Second, measure the radius from the axis to the midlines' intersection, and arbitrarily adopt this as the radius to use for that square. This procedure is defensible on two counts. First, it is simple, and in part familiar to all who have done 2-dimensional mapping.

Second, from studies long since made by the writer, the routine has surprising accuracy for finding relative ohms in the typical case. When there is doubt, subdivide in some way, treat each part as previously described, and combine.

When all radii on record, we first assign one relative ohm to each square as a temporary value; then divide unity by the several radii to find the several values of relative ohms for the squares of a tube.

A tube remainder will be a curvilinear rectangle. It will have a relative ohms value less than unity, which is first evaluated as in 2-dimensional work, and is then divided by the radius measured to the intersection of its midlines.

When all of the individual relative ohms are added, the tube's relative ohms are found. A desired equipotential line is located by counting up along the tube and by interpolation, much as in 2-dimensional work.

Quasi-3-Dimensional Maps

There is the possibility that fluid mappers can be built to solve for flow in sheets of irregular thickness, or in nonhomogeneous sheets of constant thickness, or for a combination of both variations. The mapper must be built, taking the spacing-cubed law into account. But again, once the flow pattern is obtained, mapping and analysis proceed without reference to fluid mapper flow space dimensions. Instead, the actual third dimensions and the conductivities of the real medium are used. As before, pattern lines would be arbitrarily selected to form usable tubes; and each tube would be mapped full of squares, plus a remainder. To find the relative ohms for any square, we would divide unity by the third dimension and by the conductivity for that square.

Fields with Distributed Sources

The distributed source can be simulated by sandbed types of fluid mappers. The special problems of mapping and analysis arising in such cases are beyond the scope of this paper. However, some material along this line may be found in two papers[1,2] included in the references.

References

1. Fields from Fluid Flow Mappers, A. D. Moore. *Journal of Applied Physics* (New York, N. Y.), volume 20, August 1949, pages 790–804.

2. Soap Film and Sandbed Mapper Techniques, A. D. Moore. *Journal of Applied Mechanics* (New York, N. Y.), volume 17, September 1950, pages 291–98.

3. Four Electromagnetic Propositions, with Fluid Mapper Verification, A. D. Moore. *Electrical Engineering*, volume 69, July 1950, pages 607–10.

4. The Further Development of Fluid Mappers, A. D. Moore. *AIEE Transactions*, volume 69, part II, 1950, pages 1615–24.

Discussion

John D. Kraus (Ohio State University, Columbus, Ohio): This paper presents some valuable information on field mapping in general and in particular on its application to fluid mapper patterns. The concept of the curvilinear square volume or cell is especially useful and the paper points out some of their important properties. In more detail than mentioned in the paper it is of interest to note, using the meter-kilogram second system of units, that

1. The conductance per unit axial dimension of such a volume equals the conductivity of the medium, or
2. The capacitance per unit axial dimension equals the dielectric constant, or permittivity, of the medium, or
3. The permeance per unit axial dimension equals the permeability of the medium, or
4. The characteristic impedance of a curvilinear square volume transmission line equals the intrinsic impedance of the medium.

The circling-in method described by Professor Moore for constructing curvilinear squares in a field map marks a real advance in field mapping procedures. It largely transforms what has been a free-hand art into a precise drafting technique. I have used Moore's method extensively and find that it greatly expedites the process of obtaining an accurate map.

In connection with the determination of the dimensions of a true sectoral square, bounded by two straight radial lines and two circular arcs, as in Figure 2 of the paper, it is interesting to consider the application of a curvilinear square volume transmission line to this problem. Imagine two thin metal strips running in the axial direction along the sides of a curvilinear square volume. Neglecting fringing these strips form a transmission line of characteristic impedance equal to the intrinsic impedance of the medium. If air or vacuum is the medium this is approximately 120π ohms. Now an air-filled coaxial transmission line has a characteristic impedance $Z = 138 \log (r_2/r_1)$. Furthermore, since a coaxial transmission line may be regarded as a number of parallel strip lines in parallel it follows that the larger radius of the sectoral square is

$$r_2 = r_1 \text{ anti} \log\left(\frac{\pi}{414}\theta\right) = r_1 \text{ anti } ln\, \theta_r$$

where

r_1 = smaller radius of sectoral square
θ = sector angle, degrees
θ_r = sector angle, radians
log = logarithm to base 10
ln = logarithm to base *e*, natural log

John F. H. Douglas (Marquette University, Milwaukee, Wis.): Professor Moore's fluid mappers are a real instructional aid in the teaching field. Difficulty has been encountered when the boundaries involve sharp corners and narrow slits with wire drawing and turbulence. It is to be hoped that Professor Moore will improve this part of his work and that someone will make a realistic and unbiased comparison of the electrical versus the hydrodynamic methods as to total time requirement and final accuracy.

A. D. Moore: One is grateful to Professor Kraus for mentioning additional properties of the curvilinear square volume, and for his judgment as to the usefulness of the circling-in method for obtaining curvilinear squares.

Professor Kraus's suggestion of the word "cell" to replace the longer term "curvilinear square volume" should be given a full trial.

Professor Douglas mentions having trouble with fluid mappers that have sharp corners. This is due to inertia effect. When inertia effects are not adequately eliminated by slowing the flow rates, they can be handled by going to a fluid of higher viscosity. This has already been covered by an earlier paper, see reference 4 of the paper, where the use of Methocel is discussed. Turbulence is controlled by the same measures. The narrow slit problem may or may not be handled by proper fluid mapper design; it depends on the case. Space cannot be taken here for adequate discussion of the problem.

It is to be hoped that someone will make some comparisons of time taken and accuracies achieved by alternative methods, such as graphical mapping, the electrolytic tank, use of conducting paper, and fluid mappers. However, no one would propose that one method should replace all of the others. We need all of the methods we can get, for each method has its own advantages and its disadvantages.

There is very great variety to the problems that are constantly arising. For a given problem, one method may be best; for the next, it may be the worst.

Compensating for the Quiescent Current in Multistage Magnetic Amplifiers

ALAN S. FITZGERALD
MEMBER AIEE

WHEN a single saturating reactor is used as a magnetic amplifier the output current is not zero when there is no d-c input. In the absence of any d-c saturating excitation there is a small quiescent current I_o. The application of a d-c signal causes the output to increase from this smaller value, I_o, to a larger value I_s. Under the conditions which give rise to maximum gain, the value of I_o is usually a substantial percentage of that of I_s, as for example in the order of 30 per cent.

When it is desired to use two or more such saturating reactors in a cascade multistage arrangement, it is clear that, when there is no input to the first stage, all of the succeeding stages should likewise be in a substantially unsaturated condition. Usual values of I_o usually exceed by a large margin the minimum response or threshold value of the next stage; indeed more often than not I_o is enough to drive the second stage to its full output. Thus some way must be found for annulling, or for compensating for, the effect of the quiescent current of one stage on the stage which follows it.

Several different arrangements have been devised to take care of this characteristic. For example, the circuit (A)[1] of Figure 1 employs duplicate reactors, one carrying the d-c input winding and the other being a dummy, or counterpoise, having an identical magnetic core, the same number of a-c turns, but no d-c winding. These two reactors are connected to a differential transformer so that when there is no input signal, the two opposing primary windings of the transformer receive currents equal in phase and magnitude and the secondary winding delivers zero output.

A somewhat similar arrangement (B)[2] also employs the counterpoise principle. In this case the two reactor a-c windings are arranged as two legs of a bridge circuit, which includes the midtapped secondary of a transformer, from which the amplifier output circuit is energized. In like manner, when there is no d-c input, the bridge is balanced and the current in the output circuit, such as a relay winding for example, is zero.

In another method (C)[2] a bucking winding is provided on the relay; this winding carries a fixed current adjusted so as to produce excitation equal and opposite to that which is due to the I_o in the main winding. When there is no input signal the relay receives zero ampere turns. This increases to a value proportional to $(I_s - I_o)$ when a d-c signal is applied.

The relative advantages and disadvantages of these three systems have been discussed elsewhere[3] and need not, therefore, be recapitulated here. It will, however, be evident that the amount of high-permeability high-cost core material required in the case of circuit (C) will be less than that necessary for the other two arrangements.

It is with the performance of the several types of circuit arrangement that we are now concerned; particularly in respect to their ability to develop gain at low input power levels.

The application of the input signal to the d-c winding of the saturating reactor causes a change in the impedance of the a-c winding. When the a-c winding is energized at an appropriate level, for which a definite optimum value exists, the signal results in a commensurate volt-ampere or power change in the a-c winding.

The problem is to make use of this power change in our output circuit, in which it is desired that the power be zero when there is no signal.

Not all of this power change can in every case be utilized in the output circuit. The original change of power level in the a-c winding is subject to varying degrees of loss or attenuation in the different circuit arrangements or networks. The pay-off in devising a compensated magnetic amplifier is the extent to which attenuation can be minimized.

In the arrangement (A) all of the interstage power is transferred by transformation. At low power levels the question of efficiency of transformation is a cardinal one. The circuit (B) is better than that of (A) in that the transformer used in this arrangement is energized from the source and not exclusively by interstage power.

In the arrangement (C), on the other hand, the whole of the power increment in the a-c winding is conveyed to the output device. That is to say, no loss in interstage transfer of power results from the presence of the compensating arrangement nor is there any attenuation introduced which would not exist in its absence. The power supplied to the relay compensating winding is not interstage power.

In Figure 2 gain curves are shown for each of the three arrangements, using the same identical reactor. This is a 3-leg type core of *EI-3A* mumetal stacked to a square center section.[4] The curves are based on maximalized gain computations such as the author has described elsewhere.[5] That is to say, the output circuit resistance is adjusted to the optimum value for each value of output plotted; and the a-c voltage is maintained at a constant value, which is that at which the gain is greatest at the value of input

---

Paper 52-5, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted July 27, 1951; made available for printing November 13, 1951.

ALAN S. FITZGERALD is an electrical research engineer, San Francisco, Calif.

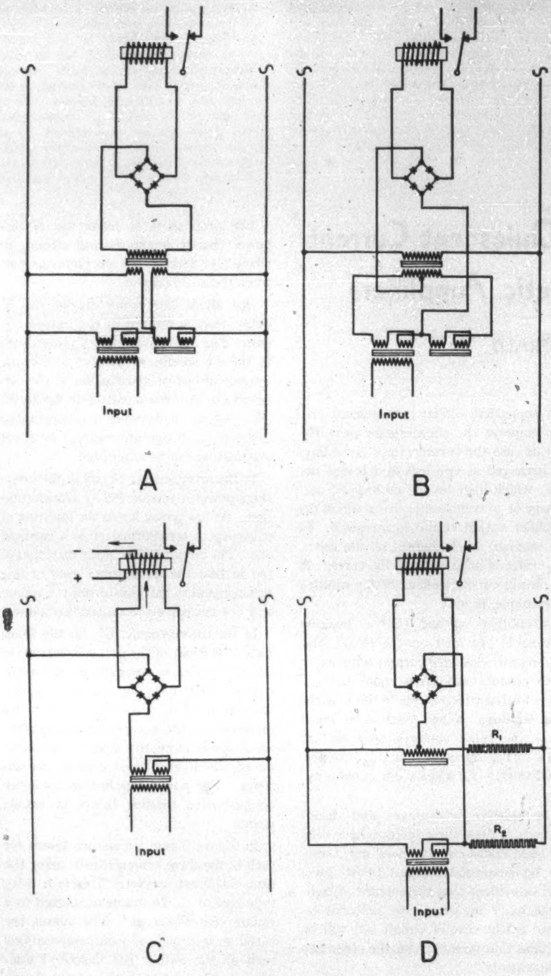


Figure 1. Four ways of compensating for quiescent current

signal at which the peak of the gain curve occurs.

In the case of arrangements (A) and (B) the output current is zero when there is no input, and the output power is the output circuit resistance multiplied by the output current squared. In the case of (C) the output current is increased from I_o to I_s and the output power is the output circuit resistance multiplied by $(I_s-I_o)^2$.

For the transformer which was used in the circuit (A), a mumetal core was employed in view of the low power levels involved and the transformer was designed to give maximum efficiency at the peak of the gain curve. In the arrangement (B) the characteristics of the transformer, since it is merely a source of power, are not a material factor in the performance of the circuit. All of the curves refer to 60 cycles.

The curves show the very definite superiority of the arrangement (C), in respect to the gain developed, over the other two circuits. The systems (A) and (B) also require larger amounts of material resulting in greater weight and cost. Thus for purposes in which the output of the magnetic amplifier is to be used for furnishing ampere turns, this method has merit. It is, for example, suitable for use in multistage magnetic amplifiers.

It is equally applicable to relays and other electromagnetic control devices. In this case, though, as compared with (A) and (B) it has the disadvantage that stock relays cannot be employed without dismantling them and reassembling them with special coils. It also is not applicable where an a-c output is required.

Clearly, a magnetic amplifier hookup that would combine the best features of these three methods is the thing to aim at. Such an arrangement should desirably:

1. Have gain and threshold of response approaching that of (C).
2. Have zero output when there is no input.
3. Require a minimum of material and not more than a single core of expensive high permeability alloy.
4. Be equally adaptable for a-c or rectified output.

The author has attempted to meet these requirements with the arrangement (D)[6] which comprises, in addition to the single saturating reactor with the usual a-c and d-c windings, a transformer and two resistance elements.

The transformer is made of regular grade transformer-type core material; no significant improvement either in gain or compensation appears to accrue from the use of nickel alloy of the same type as is used in the reactor. In the diagram the primary comprises the full number of turns and the secondary is the section to the left of the tap. There is an operational convenience in using separate windings in laboratory development, but autotransformer connections can be used in production. The number of primary and secondary turns, not the ratio merely, depend in the first place upon the characteristics of the saturating reactor and in the second place on the power level at which it is desired to operate the amplifier. A step-down ratio in the range of from two to four is representative.

The correct compensating or neutralizing action, that is to say the adjustment of the circuit constants so that the output is substantially zero when there is no input, is achieved when the transformer has the right ratio and draws the correct magnetizing current, in relation to the reactor

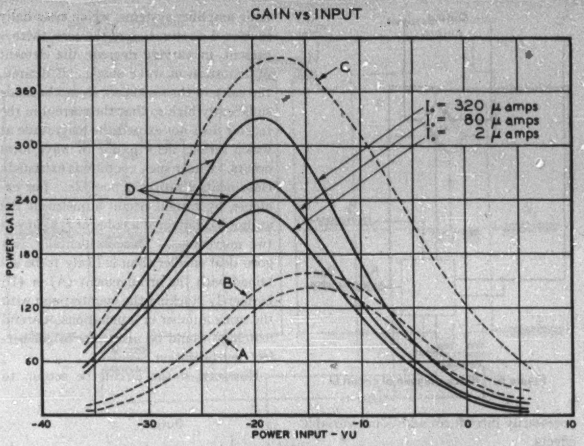


Figure 2. Comparison of gain of four compensating circuits

characteristics and the values of the two resistances, so as to bring about the proper vector relations as shown in Figure 3.

In the vector diagrams the relations indicated when there is no signal are substantially authentic. In the case of the conditions which result when a signal is applied, the diagram is believed to be reasonably suggestive of the magnitudes of the several voltages involved; but it is not possible to determine or to present the phase relations of nonsinusoidal voltages with significant precision.

The circuit operates at a voltage somewhat greater than that which would be used were the saturating reactor a-c winding and the rectifier to be energized in series directly from the a-c source. The 25-volt supply commonly used for low power control circuits is suitable. This permits the branch of the circuit which comprises the a-c winding and the resistance R_2, which has the higher value of the two resistances, to be operated under constant current conditions. It is this feature of the design of the circuit which is an important element in securing the values of gain which this system achieves.

The path through which the power from the a-c supply reaches the output is formed by the resistor R_1 and the transformer. Thus, increasing the value of R_1, decreasing the transformer secondary turns, or increasing the turns of the primary will all tend to cut the gain.

Using the identical core, a-c and d-c windings, and rectifier, with which the performance of the other three arrangements was obtained, the gain of the new circuit (D) is also plotted in Figure 2.

The gain figures shown in Figure 2 are very moderate ones in the light of figures usually encountered at the present time, for several reasons. In the first place all of the tests were made at 60 cycles. Furthermore, at the time this work was done the improved core materials having characteristics superior to mumetal were not available. And in addition, no self-saturating or feedback feature is included since it is desired to present the most direct comparison possible between the inherent basic performance of the several circuits. The inclusion of feedback

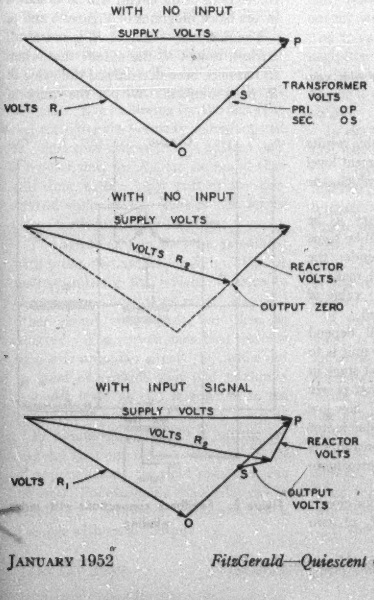


Figure 3 (left). Vector diagram of circuit D

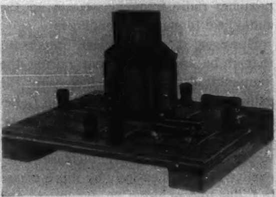


Figure 4 (right). Complete setup of circuit D

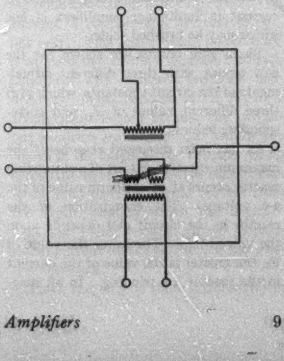


Figure 5 (right). Block diagram of saturating reactor and compensating transformer

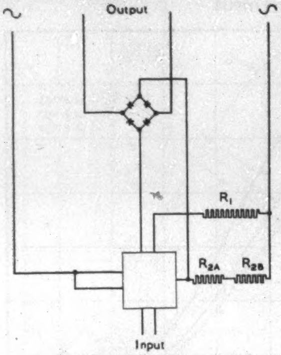


Figure 6. Block diagram of circuit D

necessarily introduces additional variable effects.

It may at first appear that the recent availability of greatly superior core materials might render the problem of compensating for the quiescent current a matter no longer of very great import. It is to be pointed out, however, that the performance of magnetic amplifiers is very much improved at frequencies greater than 60 cycles and that the improvement in core metallurgy, and increased future exploitation of magnetic amplifiers, are very definitely related to the use of higher frequencies. For example, the use of an a-c supply which approaches radio frequency opens up the audio field to magnetic amplifiers.

However, it is particularly important that it be appreciated that raising the frequency also results in increasing the values of I_o, relative to the values of I_s, because of increased core loss effects; and the extent to which it is possible to secure further gain by increased frequency is limited entirely by this increase in I_o due to core loss.

It is, therefore, not believed that this problem of compensating for the quiescent current in multistage amplifiers is one which may be brushed aside.

Three gain curves are shown for the new circuit with three different adjustments of the circuit constants, which give three different values of I_o, and corresponding values of gain.

As has been discussed elsewhere[4] the maximum effective output of a saturating reactor occurs at an optimum value of the a-c voltage. The energization of the reactor in the circuit (D) depends upon the value of R_2. The lower the value of R_2, the greater is the value of the current in the reactor a-c winding. In all magnetic amplifier systems, which essentially involve d-c saturation of the core, there is present, in varying degrees, the element of distortion of wave shape. If desired, the value of the resistance R_2 can be made sufficiently high so that the current in the reactor does not exceed the magnitude at which serious deformation of wave form occurs. Under such conditions extremely close neutralization is possible. For example, by careful circuit adjustment the author has obtained a value of I_o as low as two microamps. This, incidentally, is a good deal smaller than is likely to be attained with the arrangement (A) or (B) by merely stacking the counterpoise with the same number of laminations. Careful matching would be necessary to get perfect neutralization.

However, if the circuit be set up to achieve such a low value of I_o, this results in operating the reactor at a current level somewhat less than that at which the highest gain is possible.

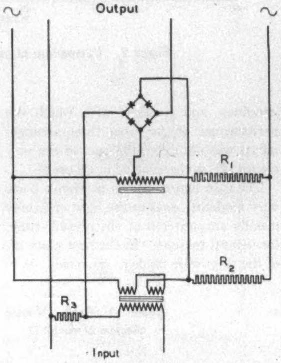


Figure 7. Feedback connections with coupling resistor

If the value of the resistance R_2 be made somewhat lower there will be some slight wave form distortion under the no-input condition, which will make it impossible to reach such a low value of I_o; but the gain will be improved.

The proper adjustment will depend upon the purpose for which the unit is to be used; as for example for what stage in an amplifier it is intended; what power level it is designed to operate at; how low we want I_o to be. The particular factor which determines the circuit constants is the threshold of response of the proximate stage.

The three curves are plotted for circuit adjustments which give an I_o of two microamperes, 80 microamperes, and 320 microamperes. The maximum gains, respectively, are 220, 260, and 330. If the ratio of the I_o to the I_s currents, with an output circuit resistance at which the maximum gain is developed, be expressed as a percentage, for the 2-microampere setting this is less than 0.1 per cent; for the 80-microampere setting it is about 2 per cent; and for the highest gain setting it is of the order of 6 per cent.

In this connection it is perhaps of interest to note that in the arrangement (C), at maximum gain, I_s is 13 milliamperes and I_o is six milliamperes.

Incidentally, the I_o current in circuit (D) is almost entirely a second harmonic derivative. This double-frequency current could undoubtedly be materially reduced by filtering if there were any advantage to be gained by so doing. However, the attractive feature of this arrangement would appear to be the economy of material.

The consumption of the network, for the three different adjustments shown in the curves, lies in the range of from one to two voltamperes.

The compensating transformer is extremely simple and inexpensive to construct, and need not be of substantial size relative to the reactor. Since the reactor and the transformer are matching units for any given circuit specification, it is logical to make a subassembly of these two elements, as shown in Figure 4. They may be mounted in a single can so that the connections of the circuit will be as shown in the block diagrams of Figures 5 and 6.

The resistance value of R_1 is not at all critical, except to the extent that when R_1 has once been determined the value of R_2 must conform. Within the range of

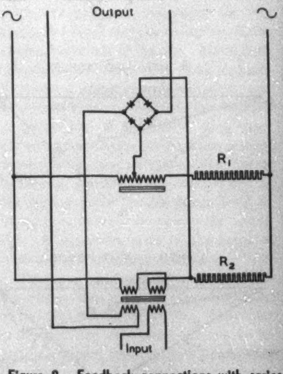


Figure 8. Feedback connections with series winding

resistance values embraced by this type of circuit most resistor manufacturers can furnish a stock vitreous enameled fixed resistor within ten per cent of any desired value; this is close enough for the value required for R_1.

In order to obtain the desired compensated or neutralized condition, a somewhat closer adjustment of R_2 is necessary. For this purpose rheostats or other slide-contact devices, though provided with effective locking gadgets, are not too attractive. The author prefers to obtain the desired value for R_2 by the use of resistors of dissimilar value R_{2a} and R_{2b} as shown in Figure 6. If R_{2a} be chosen to comprise a substantial portion of the desired resistance value, and R_{2b} be selected to make up the balance, it is usually possible without difficulty to obtain the value necessary to give the minimum value for I_o, by a suitable combination of two standard catalog items.

The values of R_1 and R_2 depend upon the characteristics of the reactor and upon the degree of neutralization desired; that is to say how close to zero it is desired to attain. R_2 is greater than R_1, a ratio in the order five times being representative.

For example, in a typical instance the indicated values for R_1 and R_2, respectively, were 2,210 and 6,430.

For R_1, values of 2,000, 2,250, and 2,500, are to be found in the resistor stock list; any one of these values will be satisfactory.

For R_{2a} there is a listing of 6,000. Values of 350, 400, and 450 are available for R_{2b} giving available values for R_2 of 6,350, 6,400, and 6,450, between which values there are differences of less than 1 per cent. Even allowing for the fact that resistor units are normally manufactured to 10 per cent tolerance, it is evident that a resistor unit for R_{2b} can be found that will come within 1 per cent of any prescribed value of R_2. This is sufficiently close for effective neutralization.

The amplifier unit may be assembled with prescribed values for R_1 and R_{2a}. A resistor unit for R_{2b}, which conforms to the tolerance deviations of R_1, R_{2a}, and other circuit components, can then be selected. The curves show that the new magnetic amplifier circuit (D), while not so good as regards gain and minimum response level as (C), is superior to the two systems.

In comparison with (C) it has the advantage that any output circuit or device of approximately the right resistance can be used. For example a stock relay can be employed. Moreover, compensation is possible with an a-c output.

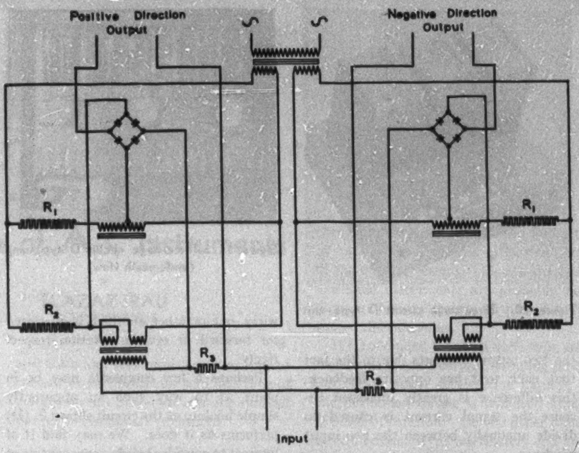


Figure 9. Push-pull polarized balance detector arrangement

Compensation is independent of balance between, or aging of, rectifiers; the compensation adjustment is therefore permanent. It is also much less affected by temperature effects and by line voltage deviation than is the case with arrangement (C).

With the (D) arrangement, in a multistage amplifier, each stage can be individually tested as a separate unit and the output read directly with a milliammeter. The I_o value is often so small it may be neglected in computing power output; it cannot usually be seen on a meter having a scale suitable for indicating the output.

In the arrangement of (C), neutralization can only be accomplished when all the stages are assembled and operating together; it is necessary to adjust the interstage currents to minimum values which are substantially removed from zero. Likewise the power output of a stage can only be computed on the basis of the difference between the I_s and I_o current values.

A further advantage of this arrangement, as compared with the circuit (C), is apparent when the addition of feedback is considered. This can readily be provided by an additional resistor R_3 connected in series with both input and output as shown in Figure 7. This is not possible without neutralization, in the absence of which there would be feedback from the residual, or I_o, current when there is no input.

If the input and output circuit resistance values are incompatible, or if it is desired to keep the input and output circuits electrically separated or insulated, a feedback winding may be used instead of a resistor, as shown in Figure 8.

Where feedback is used with a saturating reactor which is not compensated, it is necessary to provide an additional bias winding which usually calls for an extra rectifier. This is not necessary in the circuit (D).

Figure 9 is a push-pull feedback arrangement consisting of two units similar to Figure 7. The connections to the R_3 resistors are made so that the feedback is positive on one side when it is negative on the other; and vice versa. Thus the feedback action tends to produce a greater output from one side and a reduced output on the other side, in accordance with the polarity of the input signal.

However, the selective or discriminating polarized action is not limited to this amount of increase and decrease in output only. This result would be obtained if we used two entirely separate feedback amplifiers with reversed feedback connections. What happens in the Figure 9 arrangement is that because the two input windings are connected in parallel, as shown in the diagram, the voltages set up across the R_3 resistors, by the output currents, tend to reject signal current from the input winding which has negative feedback; and to suck in or increase signal current on the side which has positive feedback.

Thus, not only is there a difference in

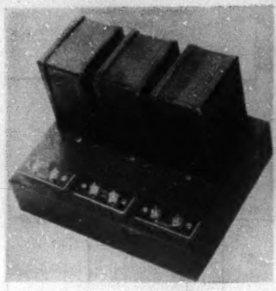


Figure 10. Three-stage circuit D type unit

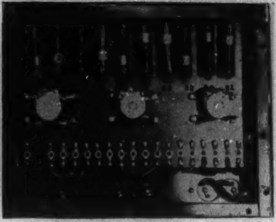


Figure 11. Three-stage circuit D type unit (underneath view)

the two output currents due to the fact that each unit has opposite feedback; this difference is greatly increased because the signal current is caused to divide unequally between the two input windings.

By means of arrangements based upon Figures 7, 8, and 9, it is possible to make sensitive polarized amplifiers in which, over wide ranges of signal level, output is delivered by one side only, the current in the other side being of negligible magnitude. When the signal polarity is reversed the outputs of the two units likewise change over.

The advantage of the arrangement which gives two separated outputs, one for a positive signal and one for a negative signal, is that, when it is desired further to amplify the output of a polarized arrangement such as Figure 9 by means of additional stages, as, for example, up to a power level suitable for driving a sizable motor, this may be done by means of simple or neutral types of amplifier such as those shown in Figure 1. These arrangements are more effective in handling substantial amounts of power than the usual polarized types of magnetic amplifier in which substantial losses usually occur in the differential or balancing circuit structure. Furthermore, most types of motors used in reversing control systems do not operate with reverse polarity input, but more usually have two different wires, one or other of which is energized for forward or reverse rotation respectively.

Perhaps a few comments may be in point as to why such an apparently simple hookup as the circuit shown in (D) performs as it does. We may find it of interest to consider briefly some aspects of bridge circuits in general.

Perhaps as a relict from school days the impression is prevalent that a bridge works best when all the four arms are of equal magnitude. This is true under restricted conditions but is ~~not to~~ be construed as a generality. Let us examine the premises. In the use of a Wheatstone bridge for measurement purposes certain postulates are accepted. One of these is that the battery voltage is assumed to be a constant. For this case the rule as to equality of the four arms may be valid.

But in a magnetic amplifier bridge circuit we are not subject to such a restriction. The premises are different. The constant term, in any of the magnetic amplifier bridge or differential circuits, is not the supply voltage but the a-c energization, in the absence of the input signal, of the saturating reactor. As we have discussed above, there is a definite value of voltage, current, or flux density at which the reactor gives the maximum value of gain for any given input signal. This is a deciding factor in determining the circuit constants of our bridge. And it happens that proportions for the arms of the bridge, departing materially from equality, give considerably more current in the galvanometer circuit of our bridge, than if the four arms are of the same value. Incidentally, it may be of interest to note that a simple bridge consisting of a duplicate core or counterpoise and two resistances of the same impedance value, comes out considerably below the performance of (A) and (B) in the curves.

Briefly, when the impedance of the saturating reactor is decreased as a result of the input saturating current, the current in the a-c winding increases. This current increment is derived from the source through two channels, one of which includes the output circuit, and one of which does not. If we can minimize the increase of current through the latter channel, by using such a high value of resistance that we approach a constant current condition; and if we can arrange that the alternate path which passes through the output circuit, presents a minimum of attenuating impedance: then more of the increased reactor current will flow through the output circuit and a higher gain will result.

A modification of the circuit arrangement (D) may be used with some types of self saturating magnetic amplifier circuits.

Figures 10 and 11 show above and below views of a typical 3-stage magnetic amplifier based on the circuit (D). These illustrations are included only to emphasize the extreme simplicity of the physical structure and the extent to which it is completely conventional.

References

1. Electrical Translating Apparatus, **F. W. Lee.** United States Patent reissue re. 19,129, April 3, 1934. (Original U. S. Patent 1,855,639, April 26, 1932.)

2. Magnetic Amplifier, **A. S. FitzGerald.** United States Patent 2,027,311, January 7, 1936.

3. Magnetic Amplifier Circuits Neutral Type, **A. S. FitzGerald.** *Journal*, Franklin Institute (Philadelphia, Pa.), volume 244, page 249.

4. Some Notes on the Design of Magnetic Amplifiers, **A. R. FitzGerald.** *Journal*, Franklin Institute, (Philadelphia, Pa.), volume 244, November 1947, figure 6, page 332.

5. Ibid., pages 324 and 325.

6. Magnetic Amplifier System, **A. S. FitzGerald.** United States Patent 2,461,046, February 8, 1949.

No Discussion

Overload Protection of A-C Instruments

WILSON PRITCHETT
MEMBER AIEE

ELAZAR TRAU
NONMEMBER AIEE

Synopsis: The behavior of indicating instruments subjected to overloads is described. The principles underlying design requirements of protective devices are presented and certain features of instrument design which enhance the effectiveness of protective devices are mentioned. Test results on current circuit and voltage circuit protective devices which are mounted within the case of a single-phase wattmeter are given. The current circuit tests cover overloads ranging up to over 200 times rated current. The latter represent direct connection of the current circuit across a 240-volt power circuit while the voltage circuit of the wattmeter was energized.

THE possibility of rendering electrical indicating instruments immune to damage due to any reasonable overload condition appeals to those responsible for maintaining an electrical laboratory. Previous work on overload protection has been limited largely to the protection of sensitive d-c instruments.[1–6] Although a-c instruments are usually more rugged than the d-c types, overloads on conventional instruments, except moderate overloads on the switchboard models, require repairs involving opening of the case. Necessary repairs range from straightening the pointer and rebalancing to cleaning or replacing the entire mechanism and rescaling. The cost of the latter often nearly equals the price of a new instrument.

Examination of instrument damage records in university laboratories has shown that current-circuit overloads due either to carelessness or accident are more common than all other types of instrument abuses combined. For this reason the principal objective of the work reported in this paper has been directed toward current circuit protection. The effects of overloads on the various types of a-c instruments are discussed and the requirements to be met by protective devices are outlined. The principles of operation and test results on one type of current circuit protector and one type of voltage circuit protector meeting most of the requirements are presented.

Mechanical Behavior During Overload

The application of a sudden overload develops a torque which accelerates the moving parts of the instrument. Mechanical damage may occur either during acceleration or during subsequent deceleration when the stop is reached. Because of the fundamental operating principles of the instruments, current-circuit overloads in wattmeters produce torque proportional to the current overload and in electrodynamic ammeters proportional to the square of the current overload. In moving-iron instruments the torque is very nearly proportional to the square of the overload current only for moderate overloads, because overloads in excess of several times rated current tend to saturate the vanes. Measurements made on some common types of iron vane ammeters have shown that the developed torque increases little after the overload exceeds about 20 times full-scale current and is limited to about 200 times full-scale torque.

During acceleration either the damping vane or the pointer or both may be deformed with respect to the torque producing element—the moving coil or the moving vane. Static load tests on the moving systems of high quality conventional portable instruments have shown that the pointer is capable of withstanding accelerating torques a little greater than 1,000 times full-scale value without suffering discernible permanent deformation. Damping vanes of the eddy-current type which rotate in a plane perpendicular to the axis are naturally even more rugged, but the air damping vane of the corner supported type has been found to be fragile when compared with other parts of the moving system. Thus many conventional instruments are probably not damaged by accelerating torques of 1,000 times full-scale value.

It is concluded that iron vane type ammeters using eddy current damping probably never suffer acceleration damage if the developed torque is balanced about the axis of rotation. On the other hand, electrodynamic-type wattmeters and ammeters may be damaged during acceleration.

If in ordinary laboratory work it is assumed that the maximum accelerating torque likely to be encountered by a portable instrument results from accidental connection of the current circuit to a 240-volt power circuit, the conditions necessary for preventing acceleration damage under usual circumstances can be established.

In the case of the wattmeter it is necessary to prevent the maximum current from exceeding 1,000 times the current required for full-scale deflection. For a 500-watt wattmeter with a 5-ampere current coil, this is about 1,500 amperes if the low-voltage terminal (150-volt) is energized or 3,000 amperes if the proper voltage terminal (300 volt) is energized. In the case of a 5-ampere electrodynamic ammeter the maximum allowable current is $\sqrt{1,000}\times 5$ or 160 amperes.

In many practical cases the overload current in the current circuit of an instrument increases almost linearly with time during the first millisecond because sufficient inductance is usually present. To prevent acceleration damage to conventional instruments protective devices

---

Paper 52-7, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted April 24, 1951; made available for printing November 13, 1951.

WILSON PRITCHETT is Assistant Professor of Electrical Engineering and ELAZAR TRAU is a graduate student, both with the University of California, Berkeley, Calif.

The suggestions and criticisms by Professor R. M. Saunders and the work of F. F. Davis and of the late Fred Lory on the construction of the protectors are gratefully acknowledged.

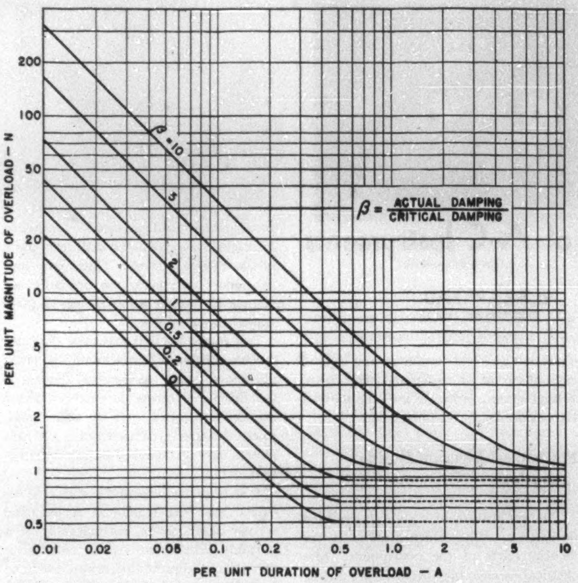


Figure 1. Conditions required to give a maximum "throw" of full-scale deflection to the moving system of an instrument

$$N=\frac{\text{Actual overload torque}}{\text{Full-scale torque of restoring spring}}$$

$$A=\frac{\text{Duration of overload}}{\text{Undamped natural perid of moving system}}$$

β = Relative damping

are required to operate before the current becomes greater than the values mentioned above. Since the inductance of these circuits is about 0.1 millihenry the protective device must operate within about 1 millisecond for the wattmeter and about 0.1 millisecond for the ammeter.

Deceleration in an overloaded instrument severely abuses the part (usually the pointer) which reaches the end of its travel first, because this part transmits the decelerating torques to all parts in addition to the overload driving torque. The moving systems of most conventional instruments are capable of withstanding only comparatively small decelerating torques. This means that effective instrument protection involves terminating the accelerating torque before sufficient kinetic energy has been acquired by the moving system to damage it at the end of its travel.

Figure 1 shows the relationship between the magnitude of the overload torque and its duration which will limit the first "throw" of the pointer to full-scale deflection (unit deflection). The data apply to any system represented by a second order linear differential equation with constant coefficients. It is assumed that the initial deflection and velocity are zero and the amplitude of the overload torque is constant during the overload interval. Data for the curve were computed from the equations given in the Appendix.[6] Other nondimensional curves applying to such systems may be found in the literature.[7,8] For completeness the data in Figure 1 are given for a wider range of the variable and the parameter than needed for the subject of this paper.

In many practical cases because the overload current initially increases linearly the final value of the overload torque, N, may be allowed to be twice that given by the straight sections of the graph. In an electrodynamic ammeter the final value may be three times that given because the torque pulse is parabolic. Although a factor of three may be applied for this condition, the possibility of acceleration damage already mentioned is the controlling factor. These multiplying factors of two and three apply only during an interval short compared to the electrical time constant of the circuit. In a concentric vane ammeter the torque pulse may be almost rectangular and the value of N given in Figure 1 should be applied directly. If the pointer of an instrument and its stop are designed to withstand an appreciable decelerating torque, the overload, N, or the duration, A, may be increased further by an appropriate factor.

Application of Figure 1 to conventional wattmeters and iron vane ammeters shows that protective devices operating within about 1 millisecond afford protection from deceleration damage. If acceleration damage can be prevented in electrodynamic ammeters, deceleration damage probably would not occur.

Figure 1 shows that increasing the relative damping or increasing the undamped natural period of the instrument allows a greater overload or a greater duration of overload. The curves have been terminated at $A=0.01$ on the left side so that a larger scale could be used to make the graph easier to read. In the straight regions of these curves the effect of extending the range to the left may be had by multiplying N by 10^n and dividing A by 10^n, where n is an integer.

Thermal Behavior During Overload

During overloads of short duration the temperature rise of conductors within the instrument is very nearly proportional to the electrical energy released because the heat dissipated during that short interval is negligible. If the overload current is assumed constant, the temperature rise in degrees Centigrade for copper wire is:

$$\theta=\frac{I^2Rt}{192} \text{ (approximately)} \qquad (1)$$

where

I = magnitude of constant current in amperes
R = ohm per pound at the initial temperature for the wire size used
t = duration of the current, I, in seconds

Only the wire size and not the resistance

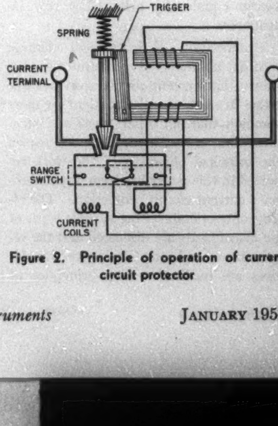


Figure 2. Principle of operation of current circuit protector

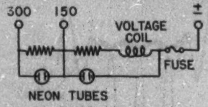


Figure 3. Circuit sketch of voltage circuit protector using fuse and neon tubes

of the instrument is needed in applying equation 1. For No. 14 gage copper wire initially at 20 degrees centigrade the temperature rise is very nearly $I^2t \times 10^{-3}$ degrees centigrade. For this size, which is commonly used in 5 ampere current coils, the most severe overload considered (3,000 amperes) would produce a temperature rise of 65 degrees centigrade in 7 milliseconds. This shows that the allowable duration is several times greater than it is in the case of protection from mechanical damage. On the other hand it is certain that moderate overloads allowed to persist result in thermal damage. This factor co-ordinated with the maximum allowable mechanical overload may be used to determine the point at which the torque shall be removed if damage is to be prevented. The authors have merely considered either full-scale current or maximum rated current as the point at which immediate protection should be applied, and none of the principles of thermal time delay devices have been employed. The only disadvantage this seems to offer is the inconvenience to the user of a wattmeter operating at low power factor where it may be desired intentionally to place a momentary overload upon the current circuit.

Instrument Protector Design Considerations

The authors have been guided by the principle that it is as improper to open the current circuit of a current-measuring instrument as it is to short-circuit the voltage circuit of a voltage-measuring instrument. For this reason the use of fuses in current circuits has not been considered in this paper. Although fuses in combination with other elements might be used the authors have been unable to adapt any of these to their satisfaction.

Features which it is felt should be incorporated in the design of protective devices for portable a-c instruments should include the following:

1. The protective device should be contained within the case of the instrument.
2. The device should not affect the accuracy or usefulness of the instrument.
3. It should protect the instrument from every reasonable overload condition.
4. The protector should be capable of being reset easily and without opening the main seal on the instrument case.
5. It should be impossible to reset the protector against an overload.
6. A conspicuous warning should be visible as soon as the protector has operated.

Protective devices having most of these features have been constructed and tested by the authors. The operating principles together with practical applications and test results now will be given.

Current Circuit Protector

Figure 2 illustrates the components and circuit connections of the current-circuit protector. The range switch of the instrument connects the coils in series for the lower range of the instrument and in parallel for the upper range. Each current coil of the instrument is permanently connected in series with one of the coils on the electromagnetic tripping device. A current overload on either range causes the trigger to be withdrawn from under the spring-loaded short-circuiting plug which snaps downward. The resulting short-circuit protects the current coils from overload. The protector is reset manually by lifting the plug and allowing the trigger to return to its original position. It is seen from the construction that static friction is the only agent preventing the protector from tripping while current flows in the current coils. Whenever the instantaneous current reaches a value sufficient to cause the trigger to overcome the static friction, the trigger suddenly moves out from under the spring-loaded short-circuiting plug. Accordingly the critical tripping current for this kind of protector is the peak value of the current wave which is just able to withdraw the trigger.

Voltage Circuit Protector

Figure 3 shows the voltage circuit protector. The purpose of the neon tubes is to permit the use of a fuse with a much higher rating than the maximum continuous current rating of the voltage circuit. Available fuses with a rating low enough to be used alone have a resistance that is both too high and too variable to be used without affecting the accuracy in the voltage circuits of conventional portable a-c instruments. If either voltage range is overloaded the neon tubes fire and

Figure 4. Current circuit protector mounted inside case of Weston 432 wattmeter

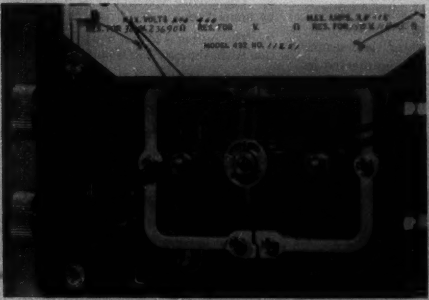
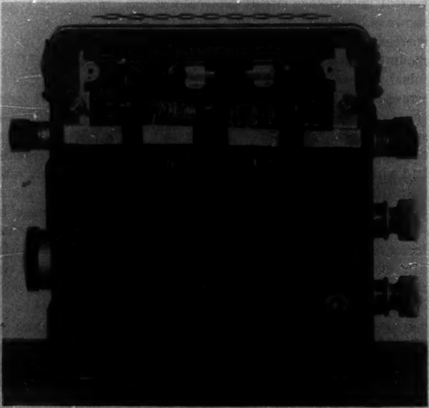


Figure 5. Voltage circuit protector mounted in multiplier compartment of Weston 432 wattmeter



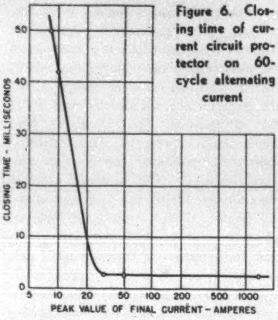


Figure 6. Closing time of current circuit protector on 60-cycle alternating current

their resultant low-voltage drop causes the fuse to blow. The critical overload voltage is the peak value of the voltage wave just able to fire the neon tubes. Although this type of voltage circuit protector is simpler and cheaper than the current circuit protector described in the previous section, its marginal utility is lower because the voltage circuits of 150-300-volt instruments are seldom overloaded. Since tubes with breakdown voltages much below 90 volts are not readily available, this type of protector is not suitable for protecting low range a-c voltmeters.

Low range a-c voltmeters are not commonly used in laboratory testing as the 75- to 300-volt instruments, but they are more susceptible to overload damage. Extension of the work reported in this paper is directed toward developing a protector of the circuit-opening type.

A Protected Wattmeter

A single-phase wattmeter having both current and voltage circuit protection is shown in Figures 4 and 5. In Figure 4 the current circuit protector is shown mounted by means of the current-terminal screws inside the case. These screws pass directly into two heavy copper electrodes which are short circuited by a spring-loaded plug whenever an overload occurs. The contacting surfaces, overlaid with silver, have a resistance of less than 10^{-4} ohm in the closed position. The small white flag on the left end of the scale uncovers a red spot when the protector trips. The screw in the lower right side of the rear of the case shown in Figure 5 is removed to admit a small dowel used for resetting the current-circuit protector. This protector was adjusted to trip at 8 amperes for the 7.5-ampere range. The impedance of the coils in the nontripped position is approximately 40 per cent of the impedance of the current coils of the wattmeter and has about the same phase angle. Figure 6 shows the closing time versus peak overload current. Figure 7 is an oscillogram of line voltage, line current, and voltage drop across the current coils. At point (A) on the oscillogram a switch was closed placing the current terminals directly across a 240-volt 60-cycle power line. At point (B) the spring loaded short-circuiting plug closed. After three cycles, point (C), the line fuse cleared the fault. The interval (A) to (B) during which the actual instrument overload occurred is seen to be a little less than 3 milliseconds. A new design of this type of protector which is under construction is expected to operate in less than one-fourth of this interval. The voltage circuit of the wattmeter was energized during the test and the kick of the pointer was hardly noticeable. This is accounted for by the fact that the line voltage was reduced to a rather low value during the fault. The test recorded in Figure 7 is typical of approximately 100 tests which have been made on the protector covering a wide range of current overloads. Following the tests the calibration of the instrument was checked and found to be unaffected. Later the instrument was placed in an undergraduate laboratory where it has been in service for several months.

The voltage circuit protector mounted in the multiplier compartment of the wattmeter is shown in Figure 5. Three neon tubes in series were used in parallel with each section of the multiplier together with a 0.0625 ampere-fuse. When the voltage circuit was operated on its 200-volt maximum range, the time required for the fuse to clear was found to be 1.5 cycles (on a 60-cycle circuit) for an overload of 7.5 per cent. For large overloads time became less than one cycle.

Conclusions

1. Current circuit overloads are the most common causes of damage to portable electrical indicating instruments.

2. On overloads greater than several times full-scale rating in unprotected instruments of the types considered, mechanical damage occurs before thermal damage and so protection from the former insures protection from the latter.

3. A current circuit protector and a voltage circuit protector for the higher voltage ranges has been constructed and tested and found suitable for protection of other a-c instruments with the exception of the electrodynamic ammeter.

4. Protection for the electrodynamic ammeter, to the degree afforded the other a-c instruments considered, is not available at this time.

5. Both protectors described are peak operated devices and their operation is therefore influenced by wave form with the result that they may trip when used on peaked waves which would not endanger the instrument.

6. It is considered that a practical solution to the problem of developing self-contained protective devices will emerge around designs in which the instrument and protector designs are coordinated and provisions for the latter are included in the design of the instrument case.

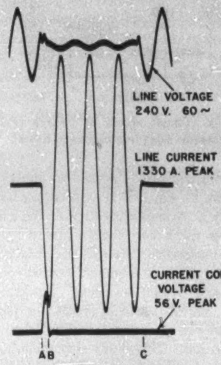


Figure 7. Oscillogram of current and voltage in overload test of current circuit protector on 240-volt 60-cycle power circuit

Appendix

The determination of the maximum allowable per-unit overload, N, which will not produce deceleration damage, as a function of the per-unit duration, A, of the overload for various conditions of relative damping, β, was made under the following:

1. The maximum swing of the moving system is taken as full-scale deflection, θ_{fs}.

2. The moving system can be represented by lumped constants.

3. The system is at rest and at zero deflection when the overload is applied.

4. The magnitude of the overload is constant during the overload period.

The following notation is used:

J = moment of inertia of the moving system
$D \doteq$ damping constant of the moving system
K = torque gradient constant of the restoring spring
θ = deflection of the moving system
$M = N(K\theta_{fs})$ = magnitude of overload torque
$U(t)$ = heaviside unit step function
$\delta(t)$ = unit impulse function
T = duration of overload
$T_n = 2\pi\sqrt{J/K}$ = undamped natural period of moving system

$$\beta=\frac{D}{2\sqrt{JK}}. \quad A=T/T_n$$

Under the assumptions given, the equation of motion for the moving system of the instrument is:

$$\frac{d^2\theta}{dt^2}+D\frac{d\theta}{dt}+K\theta=f(t) \tag{2}$$

$$f(t)=M[U(t)-U(t-T)] \tag{3}$$

$$\text{at } t=0:\ \theta=0, \text{ and } \frac{d\theta}{dt}=0$$

The results in nondimensional form are:[6]

$$N=\begin{cases}\dfrac{1}{2\sin\pi A}; & \beta=0\\[2ex] \dfrac{\sqrt{1-\beta^2}\epsilon^{ra}}{\epsilon^{2\pi\beta A}\sin\left(a+\mu-2\pi A\sqrt{1-\beta^2}\right)-\sin(a+\mu)}; & \beta<1\\[2ex] \dfrac{\epsilon^{2\pi n}}{\epsilon^{2\pi A}-1}; & \beta=1\\[2ex] \dfrac{\epsilon^{\omega p}}{\epsilon^{2\pi\beta A}\left[\cosh\left(\omega-2\pi A\sqrt{\beta^2-1}\right)+p\sinh\left(\omega-2\pi A\sqrt{\beta^2-1}\right)\right]-\left[\cosh\omega+p\sinh\omega\right]} & \beta>1\end{cases} \tag{4}$$

where

$$a=\tan^{-1}\left(\frac{\sin 2\pi A\sqrt{1-\beta^2}}{\cos 2\pi A\sqrt{1-\beta^2}-\epsilon^{-2\pi\beta A}}\right)$$

$$\mu=\tan^{-1}\frac{1}{r}; \quad n=\frac{A\epsilon^{2\pi A}}{\epsilon^{2\pi A}-1}$$

and

$$\omega=\frac{1}{2}\ \ln\ \frac{\epsilon^{2\pi A(\beta+\sqrt{\beta^2-1})}-1}{\epsilon^{2\pi A(\beta-\sqrt{\beta^2-1})}-1}$$

$$r=\frac{\beta}{\sqrt{1-\beta^2}} \quad p=\frac{\beta}{\sqrt{\beta^2-1}}$$

For $A<<1$

$$f(t)\approx MT\,\delta(t) \tag{5}$$

and the expressions for N are simplified to:

$$N=\begin{cases}\dfrac{\epsilon^{r\tan^{-1}\frac{1}{r}}}{2\pi A}. & 0<\beta<1\\[2ex] \dfrac{\epsilon}{2\pi A}; & \beta=1\\[2ex] \dfrac{\epsilon^{p\tanh^{-1}\frac{1}{p}}}{2\pi A}; & \beta>1\end{cases} \tag{6}$$

Equation 4 was used to compute the curved portions of the curves in Figure 2, and equation 6 was used to compute the straight line portions of these same curves. These curves describe the ballistic behavior of systems which can be represented by equation 2.

References

1. When Meters Blow, L. J. Fohr, Jr. *Electronics* (New York, N. Y.), January 1937, pages 22–23.

2. Protection of the High Voltage Plate Ammeter, W. E. Grove. *Electronics* (New York, N. Y.), February 1938, page 38.

3. United States Patent Number 2,451,953, C. S. Ingram. October, 1948.

4. A Sensitive Circuit Breaker for Instrument Protection, A. Borup. *Electronic Engineering* (London, England), January 1948, pages 26–27.

5. United States Patent Number 2,009,973, C. F. Nelson. July 30, 1935.

6. Overload Protection of A-C Indicating Instruments, Elazar Trau. *Thesis*, University of California, Berkeley, Calif., 1951.

7. Frequency Response Characteristics of Recording Instruments, T. D. Graybeal. *AIEE Transactions*, volume 67, 1948, pages 755–66.

8. Principles of Servomechanisms (book), G. S. Brown, D. P. Campbell. John Wiley and Sons, New York, N. Y., 1948, page 54.

No Discussion

A New Principle Is Employed for 60-Cycle A-C Network Analyzers

E. B. PHILLIPS
ASSOCIATE MEMBER AIEE

Synopsis: Through the use of a new principle, the impedance-phase-shift, the weight of the reactor elements of a 60-cycle a-c network analyzer has been reduced to a value lower than that obtainable in designs employing medium-frequency 500-cycle zone, and approximately equal to those employing high-frequency 10,000-cycle zone. The reactors have an $X-R$ ratio of infinity in terms of system measurements. Among other advantages gained by the use of standard, power frequency is the elimination of the special-type motor-generator unit to supply power to the analyzer. Power is supplied through constant-voltage magnetic-stabilizing type transformers.

THE forerunner of a-c network analyzers, the unit installed at the Massachusetts Institute of Technology in 1929, uses 60-cycle power for its operation. In comparison with those of more recent design the reactors of this analyzer are quite bulky and of considerable weight. The weight reduction of the later designs has been accomplished by: (1) the use of high-nickel-content core material; (2) a reduction of the power level at which the analyzers operate; (3) the use of frequencies higher than 60 cycles or the use of 60-cycle frequency together with the impedance-phase-shift principle. With the exception of two installations, all analyzers built since the one at Massachusetts Institute of Technology operate at frequencies higher than 60 cycles.

One of these 60-cycle analyzers,* purchased by the University of Kansas and placed in use in July 1947, uses the impedance-phase-shift[1] principle. An additional improvement has been made recently in the further application of this principle. This paper describes this latest form of the basic principle and a few of the exclusive features made possible by its use.

Requirements of Analyzer Reactors

Before proceeding directly into the theory of the principle it may be of interest to the reader to review the general requirements for analyzer reactors. The reactors are used to obtain the inductive reactance component of the impedance units used for the representation of transmission lines, transformers, and synchronous machines and also to consume the reactive volt-amperes (vars) in the load units. For the obtainment of the wide ranges required of each of these groups, two methods of reactor construction are in general use. One method employs relatively large size tapped reactors having coils of 10 sections; the reactors being permanently connected in series arrangement. A 10-position, rotary tap switch is provided for each reactor. By the positioning of these switches various number of turns may be selected in each of two or three reactors and thus obtain reactance values having wide limits. Another arrangement uses 8 to 12 smaller reactors, each of single-section coils. By the positioning of rotary tap switches various numbers of these units may be placed in series arrangement in the line unit, parallel arrangement for the load units, to obtain the desired reactance. Either construction type may be used in connection with the impedance-phase-shift principle; the latter type, however, will be used to illustrate its application.

For either type construction and for any amount of reactance in use, it is essential that the reactors maintain a nearly constant-value inductance, usually within ±2 per cent, for a current range generally of the order of 50 to 1. It also is essential that they have a low value X-R ratio (Q); approximately 25 for those to be used for line units and 50 for those to represent transformers and synchronous machines.

Reactor Design Considerations

Reactor design formulas indicate that the weight of the active material of a laminated-core-type reactor increases rapidly for an increase in the Q requirement; approximately as the 3.1 exponential power of Q. In Figure 1 is plotted a family of curves showing the weight variation versus change in Q for a 50-ohm air-gap-type reactor having a high nickel-content, number 29 gauge laminated core. For each of the two frequencies considered, 12 reactors having various Q values within the range shown, were designed to maintain a 50-ohm reactance, to within 2 per cent, for a current range of 5 to 250 milliamperes. The core dimensions were maintained proportionate for all 12 designs of a given frequency in order that discontinuities would not occur in the curves as a consequence of only certain commercial core shapes being available. Furthermore, in each design the optimum condition for the obtainment of minimum weight, that of making the core loss and the copper loss equal, has been observed except in those few instances wherein to do so would require flux densities beyond the working range of the core material. Although these curves apply only to a particular-value reactor, nevertheless the weight increase for an increase in Q is substantially the same for reactors of any ohmic value within the general range required in analyzers and employing this type construction. The design formulas also show that the weight

---

* Philips A-C System Analyzer, manufactured by the Nelson Electric Manufacturing Company, Tulsa, Okla.

Paper 52-19, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted May 15, 1951; made available for printing November 23, 1951.

E. B. PHILLIPS is with the University of Kansas, Lawrence, Kan.

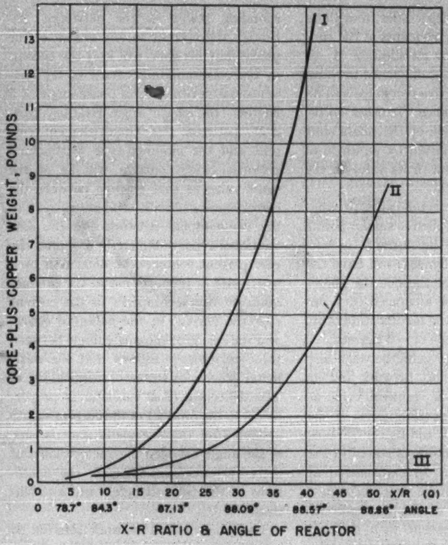


Figure 1. Variation in the weight of a 50-ohm reactor designed to have various X-R ratios. Curve I, 60-cycle designs; II, 500-cycle designs; III, 60-cycle designs for use in 5-degree impedance-phase-shift analyzers

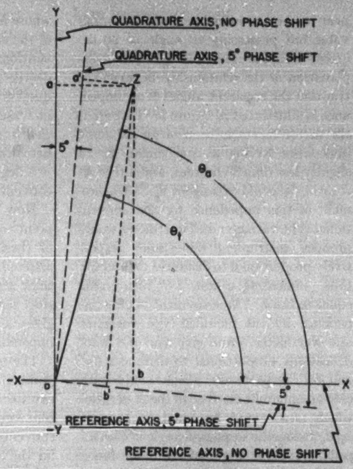

Y
QUADRATURE AXIS, NO PHASE SHIFT
QUADRATURE AXIS, 5° PHASE SHIFT
a
a'
Z
5°
θ_a
θ_t
-X
o
b
b'
5°
X
-Y
REFERENCE AXIS, 5° PHASE SHIFT
REFERENCE AXIS, NO PHASE SHIFT

Figure 2. The X-R ratio of an inductive impedance as measured by a conventional type wattmeter and varmeter, tangent θ_t, and as measured by modified type wattmeter and varmeter, tangent θ_a

of a reactor designed to operate over a given current range and to have a specified Q is very nearly constant for a wide range of ohmic values, with the exception of reactors employing compressed, powdered cores. Table I gives the total weights of reactors comprised in a group that would be used for line impedance and another for generator impedance representation.

The weight of the reactor group used for generator impedance representation on the 500-cycle design appears to be excessive. This is the actual weight, however, of this group for the condition that each individual reactor is to have a Q of 50 and each having a core of number 29 gauge laminations. A reduction of approximately 60 per cent in weight is accomplished by designing the lower value units for a Q of 25 and by the use of 5-mil thickness laminations for the cores of all. A like reduction, however, can be effected by similar design changes in the groups of the other designs, thereby maintaining approximately the same group-weight ratio for the several methods shown.

Referring again to Figure 1, curve III, the 5-degree phase-shift design, is a graph constructed from that applying to the 60-cycle design. It is obtained by moving any point on curve I, that has an abscissa value less than 84 degrees, to the right, plotting it to a new abscissa value equal to its former value increased by 5 degrees. For example, the point of curve I having a Q value of 7.5 and a corresponding angle-value of 82.4 degrees gives the point on curve III of like ordinate value but having an angle-value abscissa of 87.4 and a corresponding Q value of 22; the tangent of 87.4 degrees. Only a very short portion of curve I is required to construct the curve III.

The Impedance-Phase-Shift Principle

The basic underlying principle used in the impedance-phase-shift method is that of making an angular shift of the reference axes, the X and Y axes, as considered principally with reference to vector impedances and voltampere measurements in a-c network analyzers. Such a shift is accomplished in the instrument circuits by modifying the potential-circuit impedance of the wattmeter and of the varmeter. An angular shift of like amount is effected in each analyzer network impedance, except those having a leading angle of 90 degrees, by the use of component units each of which has been built to have impedance angles of 85 degrees instead of 90 degrees and −5 degrees instead of 0 degrees.

The potential circuit impedance of the wattmeter is altered by the inclusion in it of a small amount of capacitance and a slight reduction of its resistance, these values being such that its scalar value is unchanged but its angle changed from zero degrees to a leading angle. In a somewhat similar manner the impedance angle of the varmeter's potential circuit is changed from 90 degrees to less than 90 degrees by the inclusion of a small amount of resistance in series with an inductive reactance; the values being such as not to alter the scalar value of its impedance. For an impedance shift of 5 degrees the wattmeter's potential-circuit current will lead the applied voltage by 5 degrees while that of the varmeter's potential circuit will lag its applied voltage by 85 degrees.

Measurements of a noninductive resistor made by an ammeter, a wattmeter, and varmeter, modified as described in preceding paragraphs, would indicate an impedance of like scalar value but possessing a lagging angle of 5 degrees. But of more importance, measurements of an inductive impedance having an angle of 85 degrees, an X-R ratio of 11.43, would appear to be an impedance of like scalar value but possessing an angle of 90 degrees; an apparent infinite Q. Vector explanation of the relationship between the true and the apparent angles of an impedance is illustrated in Figure 2. Line OZ is the vector representation of an inductive impedance having a true angle, θ_t, of slightly less than 85 degrees, and a true X-R ratio, Q_t, equal to tangent θ_t. Measurement of this impedance by the conventional type varmeter and wattmeter would produce instrument deflections respectively proportional to distances Oa and Ob, thus indicating a true X-R ratio value equal to tan θ_t. Measurement of this impedance by the modified-type varmeter and wattmeter would give var and watt deflections proportional to distances Oa' and Ob', respectively, which would indicate an apparent increase in the X-R ratio, which now would be equal to tangent θ_a, θ_a being the apparent angle of the impedance. The angle of the impedance would appear to have been shifted by 5 degrees. By a similar explanation, a resistor and a capacitor in combination and having an impedance angle of -5 degrees would appear to be a pure resistor when measured on the modified instruments.

A Conventional Network and Its Equivalent Phase-Shift-Impedance Network

Consider a network consisting of shunt- and series-arranged impedances that are limited to elements having impedance angles within the range of 0 to 90 degrees; impedances that would correspond to transmission lines and loads of lagging power factors. Consider a second network, later referred to as the phase-shift network, connected in exact detail as the first, consisting of circuit elements having the same scalar values as, but whose impedance angles differ from those of the corresponding elements of the conventional network by an addition of -5 degrees; a change accomplished by a reduction in the inductive reactance and an increase in the resistance. In other words, each impedance angle of the phase-shift network has been shifted by an angle of -5 degrees from that of its value in the conventional network.

Now when equal vector-value, driving-point voltages are applied to similar points of these networks, the currents in corresponding circuit elements will have the same scalar value but those in the phase-shift network will be advanced by 5 degrees, as a result of the modification of its impedance by an angle of -5 degrees.

The vector-voltage value between any point of the phase-shift network and its neutral bus will be of identical value as that between corresponding point of the conventional network and its neutral bus, for the operating conditions stated in the previous paragraph. In explanation of this: the vector voltage drop through a line-circuit element of the conventional network that is adjacent to a driving-point voltage is $\boldsymbol{I}_1 \times \boldsymbol{Z}_1$, where $\boldsymbol{I}_1$ is the current in, and $\boldsymbol{Z}_1$ the impedance of the circuit element. The voltage drop in the corresponding element of the phase-shift network is $(\boldsymbol{I}_1\underline{/5^\circ})\ (\boldsymbol{Z}_1\underline{/-5^\circ})$, which reduces to $\boldsymbol{I}_1 \times \boldsymbol{Z}_1$. With equal vector-voltage drops existing in the line-circuit elements the vector voltages between the busses on which these lines terminate and their respective neutral busses will be of identical value. In like manner it can be shown that equal vector voltages exist between the next pair of busses and the neutrals of the two networks, and so on throughout the entire networks.

Wattmeter and varmeter readings observed on the instruments provided for each type network will be of identical respective values when made at corresponding points in the networks. Although the current in any circuit of the phase-shift network will lead the current in the corresponding circuit of the conventional network by a phase-angle of 5 degrees, the phase angle relation, however, between this circuit-element current and the modified-type wattmeter's potential-circuit current will be of the same value as that existing between the corresponding circuit-element current of the conventional network and the conventional-type wattmeter's potential-circuit current. Since the deflection of a wattmeter is proportional to the product quantity of the amperes in the current coil, the current in the potential circuit and the cosine of the phase angle between these currents, it follows that the wattmeter deflections will be of identical value. The deflection of a varmeter is proportional to the product of the current values as stated for the wattmeter and the sine of the angle between the vector values of these currents, and likewise, the deflections of the two varmeters will be identical.

The conventional network referred to in the preceding paragraphs may be likened to the network of a power system, the impedances being expressed on a per unit base. The phase-shift network may be likened to the network set up in a phase-shift-type analyzer for the representation of the system network, and whose impedances, in per-unit terms, are of the same value as the corresponding system impedances.

Load Units

As described in the preceding paragraphs all impedances set up on the network of the impedance-phase-shift analyzer, with the exception of shunt capacitor impedances, have an impedance angle shift of -5 degrees from that of

Table I. Comparative Weights of Reactor Groups Contained in Line Impedance Units and Generator Impedance Units as Effected by the Use of Three Types of Designs

Designs Employing 500-Cycle Frequency; 10,000-Cycle Frequency and Designs Employing 60-Cycle Frequency and the 5-Degree, Impedance-Phase-Shift Principle

	Line Impedance Units, Q = 25				Generator Impedance Units, Q = 50						Q = Infinity	
	500 Cycles		50 Ph. Shift		500 Cycles		10,000 Cycles		50 Ph. Shift		50 Ph. Shift	
	Core	Copper	Core	Copper	Core	Copper	Core	Copper	Core	Copper	Core	Copper
Group A*	1.840	1.803	0.394	0.384	16.38	16.02	1.637	0.267	0.514	0.506	0.828	0.810
Group B**	1.995	1.956	0.250	0.244	17.86	17.47	0.340	0.112	0.308	0.300	0.447	0.438
Group C†	2.179	2.133	0.142	0.139	19.37	18.98	0.058	0.045	0.232	0.227	0.435	0.425
Total lbs. 12 units	6.01	5.89	0.79	0.77	53.6	52.5	2.03	0.42	1.05	1.03	1.71	1.67
Total, core plus copper, pounds	11.9		1.56		106.1		2.45		2.08		3.38	

* Group A consists of three 20-ohm and one 10-ohm reactor units.

** Group B consists of one 5-ohm, two 2-ohm and one 1-ohm reactor units.

† Group C consists of one 0.5-ohm, two 0.2-ohm and one 0.1-ohm reactor units.

their system network counterparts. The apparatus contained in a load unit of this analyzer is constructed in such manner that this shift is automatically accomplished. Also, the proper amount of resistance and reactance is provided automatically in the circuit when the switch knobs of the load unit are set to the per cent kilowatt and kilovars as exist in the system load being represented.

A load-unit circuit consists of two principal branches permanently connected in parallel; one for consuming vars and the other for consuming watts. The var-consuming branch consists of eight 85-degree inductive impedances whose circuits are so arranged in relation to two, rotary tap switches that any number of them from one to eight may be connected in parallel by positioning the tap switches. One of the tap switches, the fine-adjusting one, can connect in the units designated as the 1, 2, 3, and 4 per cent kilovar units. The tap switches of the load units are of the parallel, decade type, the fine one providing for system per cent kilovar settings of 0 to and including 10 per cent in 1 per cent steps. The coarse-adjusting tap switch, wired in the same manner as the fine-adjusting one, controls the circuits to the 10, 20, 30, and 40 per cent kilovar units and provides for system per cent kilovar settings of 0 to and including 100 per cent in 10 per cent steps. A ninth inductive impedance of 0.5 per cent kilovar rating can be introduced into and removed from the load-unit circuit by a toggle-type switch. The above units provide a range of 0 to 110.5 per cent system kilovars in 0.5 per cent steps.

The watt-consuming branch of a load unit consists of eight 5-degree capacitive impedances whose circuits are so arranged in relation to two rotary tap switches that any number of them from one to eight may be connected in parallel by positioning the switches. One of these switches, the fine-adjusting one, can connect in the units designated as 1, 2, 3, and 4 per cent kilowatt units. The fine-adjusting tap switch provides for system per cent kilowatt settings of 0 to and including 10 per cent in 1 per cent steps. The coarse-adjusting switch controls the circuits to the 10, 20, 30, and 40 per cent capacitive impedances and provides for system per cent kilowatt settings of 0 to and including 100 per cent in 10 per cent steps. A ninth capacitive impedance of 0.5 per cent watt rating can be introduced into and removed from the load-unit circuit by a toggle-type switch. The above units provide a range of 0 to and including 110.5 per cent system kilowatts in 0.5 per cent steps.

Each of the nine 85-degree inductive-impedance units consists of a reactor of the proper ohmic rating permanently connected in series with a low-value trimmer-type resistor. The reactor is built to have an *X-R* ratio slightly greater than that corresponding to an impedance angle of 85 degrees. The calibration of this impedance unit consists in adjusting the air gap to give the designated ohms reactance and adjusting the trimmer resistor to obtain an over-all impedance angle of 85 degrees. In like manner, each of the nine inductive impedances are made to have impedance angles of exactly 85 degrees. The currents admitted by any parallel combination of these units will lag the voltage on the load unit by 85 degrees. However, when this current passes through the modified wattmeter it will act as though it were a 90-degree lagging current since no deflection will occur. The modified varmeter will show a deflection proportional to the product of the current through and the voltage applied to the load unit; a deflection for an applied voltage of 100 per cent, equal to the setting of the kilovar knobs.

Each of the nine 5-degree capacitive-impedance units consists of a resistor of the proper ohmic rating permanently connected in either series or parallel arrangement with a capacitor. The calibration of this impedance unit consists in adjusting the resistor to the designated value and adjusting the capacitor unit to obtain an over-all impedance angle of 5 degrees leading. In like manner each of the nine capacitive impedances is made to have an impedance angle of exactly 5 degrees. The currents admitted by any parallel combination of these units will lead the voltage on the load unit by 5 degrees. However, when this current passes through the modified varmeter it will act as though it were in phase with the load voltage since no deflection will occur. The modified wattmeter will show a deflection proportional to the product of the current through and the voltage applied to the load unit; a deflection, for an applied voltage of 100 per cent, equal to the setting of the kilowatt knobs.

A fifth rotary tap switch is provided, its scale marked in per cent bus volts having a minimum value of 90 per cent to a top value of 110 per cent in 1 per cent steps. The setting of this switch to the per cent voltage available on the load bus results in the load unit consuming the values designated by its switch settings for any bus voltage between the limits stated. To facilitate in this adjustment each load unit is provided with an individual voltmeter permanently connected to its bus.

Line Impedance Units

A line-impedance-unit circuit consists of two principal elements permanently connected in series arrangement; one of these contains 11 85-degree inductive impedance units and the other contains six 5-degree capacitive impedances and five resistor units. The individual units of each main element are designated 0.1, 0.2, 0.3, 0.4, 1, 2, 3, 4, 10, 20, and 30 per cent ohms. Three rotary-type tap switches are provided for setting the per cent ohms reactance and three for setting the per cent ohms resistance. The series-decade principle of wiring is used in connecting the switches to the resistor and reactor units thus providing a range of 0 to 71 per cent ohms system reactance and 0.2 to 71.2 per cent ohms system resistance, each adjustable in 0.1 per cent steps.

Each reactor of the reactance group has a low-value resistor permanently connected in series and adjusted during the calibration of the reactor to obtain an over-all angle of 85 degrees. Thus, when the reactor knobs are set to a desired value, say 35.5 per cent reactance, the circuit will contain three impedance units, a 30, a 5, and a 0.5, in series relation, each one having an angle of 85 degrees.

Each of six resistors of the resistance group has a capacitance permanently connected in parallel and adjusted by the use of trimmer capacitors during the calibration of the unit, to obtain an over-all impedance angle of 5 degrees. The five lowest-value resistor units are not modified to an impedance of 5 degrees leading, that is to say, they are not paralleled by capacitor units. The nonproviding of these units with capacitors introduces a negligible error in the representation of a system line which requires their use. As an example, in those instances where all five of the nonshifted units are required, a total value of 1.9 per cent ohms, the resultant inductive reactance ohms will be in excess of the correct value by 0.17 per cent ohms. Should it be desirable, in any event, to represent accurately important short lines, it may be done by reducing the fine-adjusting, reactor knob setting by an amount equal to 1/12 the sum of the knob settings of the noncompensated, resistor units being used; a reduction of either 0.1 or 0.2 per cent.

Excepting the occasional instance when this accuracy-refinement may be desirable, the setting of the knobs of a line impedance unit will be to the identical values as those for the per cent resistance and reactance of the system line repre-

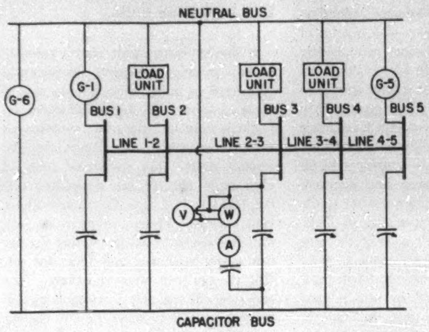


Figure 3. Illustrating the method of connecting capacitors of an impedance - phase - shift analyzer

sented. Since the per cent reactance component of a line impedance unit is composed of 85-degree impedances having a magnitudinal sum equal to the 90-degree component of the system line *Z* and since the resistor component is composed of 5-degree leading angle impedances, whose magnitudinal sum is equal to the resistance component of the system line *Z*, the analyzer impedance will be of the same magnitude as that of the line but shifted by −5 degrees.

Generator Impedance Units

Generator and other system impedances of low watt losses are readily represented on the analyzer by line impedance units. Since the reactance component of line impedance units has a zero watt loss, as viewed from the system network, it is possible to represent system impedances having extremely low resistance values; zero resistance value should occasion arise. By paralleling two line units generator ratings up to 1,000 per cent base kilovolt-amperes may be represented. Small generating stations requiring impedances greater than the range of a single line unit are accomodated by two line units in series. A single line unit, however, meets the requirements of about 85 per cent of the generating stations.

Capacitor Units

A capacitor unit consists of two capacitor groups of the decade-arranged type. Two rotary-type tap switches are used for obtaining the desired amount of capacitance, one of coarse steps having a range of 0 to 50 per cent in 5 per cent steps and a second one of fine steps having a range of 0 to 5 per cent in 0.5 per cent steps.

An inherent characteristic that follows the use of a low-frequency supply to an analyzer network, is that for a given analyzer volt-ampere base the ratio of the voltage base to the ampere base can be greater for the low-frequency analyzer than may be used for an analyzer employing a medium or a high frequency. This is due mainly to the reduced effect, almost complete absence, of self-capacitance as exists between the conductors of each circuit pair from the cord-plug cabinets and the apparatus cabinets, and to a lesser extent, the negligible amount of mutual-capacitance coupling of those portions of these circuits that, from a practical consideration, can not be shielded 100 per cent.

Thus, in general, a low-frequency analyzer may employ a higher-value impedance base than that used for analyzers of higher frequencies. The impedance base of this analyzer is 4,000 ohms, a 100-volt voltage base, and a 25-milliampere current base. As a result of the use of a high-value impedance base, the physical size of a capacitor unit required to represent a given system-capacitance effect is proportionately reduced. The largest capacitor of the adjustable unit, the 20 per cent mho step, requires but 0.133 microfarads. A fixed-capacitor unit of 500 per cent mhos requires only 3.32 microfarads.

Capacitance Representation

Advancement of the currents in the various shunt-capacitance circuits is ac-

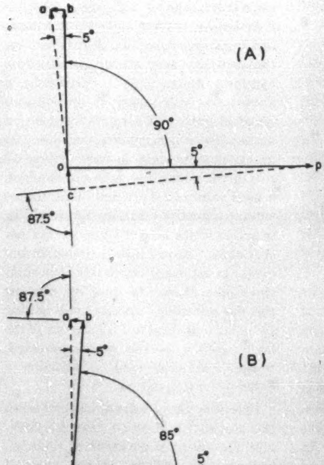
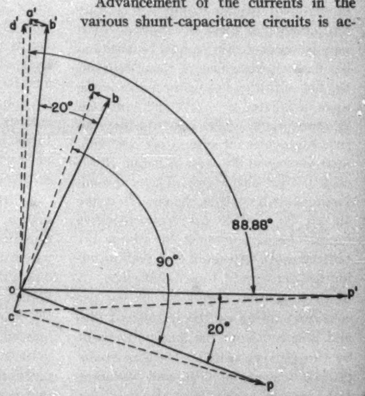


Figure 4 (left). (A) Current components of and voltage components applied to the reference capacitor, as they exist in the analyzer network. The capacitor current, I_{oa}, leads the applied voltage, V_{op}, by 95 degrees. (B) The capacitor current I_{oa} as located by measurements made on the modified wattmeter and varmeter, and its vector position in the system network in relation to the voltage V_{op}

Figure 5 (right). Accuracy of capacitance representation. $I_{d'a'}$ represents the discrepancy current of a capacitor attached to a bus whose voltage leads by an angle of 20 degrees and is 20 per cent greater than the voltage applied to the reference capacitor



$I_{d'a'}$ = (B in % mhos) x 0.000375 $\underline{/50.25^\circ}$, pu base amperes complished by the advancement of the voltages applied to their circuits. Figure 3 shows the manner of connecting the capacitor units to obtain an advancement of their applied voltages by substantially 5 degrees beyond the voltages existing between their respective busses and the neutral bus. The diagram shows the return current of the capacitors is by way of an additional bus, the capacitor bus, then from this bus through a generator unit *G-6*, the capacitor generator, to the neutral bus. The potential of the capacitor bus differs from that of the neutral bus by the amount of the vector voltage generated by the capacitor generator. The magnitude and phase position of this voltage is indicated by the line *co* in Figure 4(A).

The correct vector-value of the capacitor-generator voltage is obtained as follows: The voltage and phase-angle controls of the capacitor generator are adjusted simultaneously so that for a selected capacitor unit the per-unit current through it will be equal to the product of its per-unit mho capacitance and the per-unit value of the bus voltage to which it is connected, and further, that zero watts passes through it.

To facilitate the above adjustment, one capacitor unit, that shown connected to bus number three in Figure 3, contains a fixed capacitance branch of 5 per cent and an adjustable branch having a range of 0 to 50 per cent in 0.5 per cent steps. Permanently connected to the fixed branch is a wattmeter of the modified type, a 0-to-150-division ammeter that is calibrated to deflect 100 divisions on 5 per cent base amperes and a 0-to-150-division voltmeter calibrated to indicate 100 for an applied voltage of 100 per cent base volts. This capacitor unit, referred to as the reference capacitor, is connected into the analyzer network at the point requiring the greatest-value shunt capacitance.

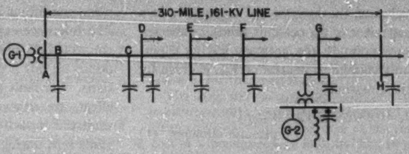


Figure 6. One-line diagram of the system network used in determining the discrepancy currents tabulated in Table II

The vector-voltage adjustment of the capacitor generator is correct when this continuously-indicating set of instruments shows wattmeter deflection of zero and an ammeter indication equal to that of the voltmeter; an adjustment that requires from 5 to 15 seconds to make and needs no correcting until a change in operating conditions is required.

The vectors representing the current and voltage components applying to the reference capacitor circuit are shown in Figure 4(A). Vector *op* represents the bus-to-neutral voltage of the bus to which the reference capacitor is connected. To this voltage is added the capacitor-generator voltage, vector *co*, resulting in the equivalent voltage, vector *cp*, which voltage is that applied to the reference capacitor. The voltage *cp*, equal in magnitude to the voltage *op* but leading it by 5 degrees, sends a current, vector *oa*, into the capacitor. This current leads the voltage of the bus to which the capacitor is connected by 95 degrees. The current I_{oa} consists of two components, I_{ob} which is due to the voltage V_{op}, and I_{ba} which is due to the voltage V_{co}, each current leading its respective voltage by 90 degrees. The current vectors in Figure 4(A) apply to the currents as they exist in the analyzer circuits. The corresponding current vectors shown in Figure 4(B) apply to these currents as they exist in the system network and as located in phase position by readings observed on the modified watt-varmeter of the master instruments.

A second capacitor unit connected to another bus having a voltage $V_{op'}$ which is assumed to be $(1.2\underline{/20^\circ})V_{op}$, V_{op} being the voltage-to-neutral of the bus to which the reference capacitor is connected, has a resultant voltage V_{cp}' (refer to Figure 5) applied to it. This voltage consists of the components V_{op}' and V_{co}, the former being the bus-to-neutral voltage and the latter being the voltage of the capacitor generator. To simplify the explanation, the size of the second capacitor is equal to that of the reference capacitor and the current vectors are those in the system network. Since the voltage $V_{op'}$ is 1.2 times the voltage V_{op} and leads it by 20 degrees, the current component $I_{ob'}$ which is due to this voltage will be 1.2 times greater than and lead by 20 degrees the corresponding current component, I_{ob}, of the reference capacitor. Moreover, this current component will be of correct value for a capacitor of any size connected to any network bus. The component $I_{b'a'}$ while proportional to the size of the capacitor is unaffected by the magnitude and the phase angle of the bus voltage $V_{op'}$. For any capacitor it will have the same phase position, that of leading the capacitor-generator voltage V_{oc} by 85 degrees, but will vary in magnitude directly with the capacitor's size. The correct vector value of the current $I_{b'a'}$ is indicated by the dash-line vector $I_{b'd'}$. This latter vector is 20 per cent greater than and leads by 20 degrees the vector $I_{b'a'}$. The current discrepancy is the vector $I_{d'a'}$, which will be referred to as an incremental current.

For a 5-degree phase shift analyzer, the

Table II. Accuracy of Capacitance Representation on the 5-Degree, Impedance-Phase-Shift Analyzer for Three Circuit Operation Conditions*

System Element, Figure 6	Heavy Load Power Flow, Mw. (1)	Heavy Load Power Flow, Mvar. (2)	Heavy Load Condition Correct % Amps.** (3)	Heavy Load Condition Deviation† %Amps. (4)	Heavy Load Condition Deviation† Deg. (5)	Light Load Condition Correct % Amps. (6)	Light Load Condition Deviation % Amps. (7)	Light Load Condition Deviation Deg. (8)	Fault Condition Correct % Amps. (9)	Fault Condition Deviation % Amps. (10)	Fault Condition Deviation Deg. (11)
Generator 1	117.0	+23.25	216	0.06	0.23	57.0	0.08	0.08	320	0.32	0.19
CD	111.6	− 0.90	218	0.04	0.01	52.0	0.01	0.02	326	0.03	0.02
Line DE at D	68.7	−15.45	134	0.05	0.06	21.2	0.01	0.08	319	0.00	0.00
Line EF at E	51.0	−20.55	108	0.08	0.17	10.7	0.01	0.18	358	0.00	0.00
Line FG at F	28.9	−30.00	68	0.19	0.34	7.3	0.03	0.22	363	0.11	0.06
Line GH at G	38.2	+30.20	86	0.01	0.03	19.1	0.01	0.03			
Generator 2	40.0	+24.80	87	0.34	0.16	25.1	0.02	0.06	400	0.68	0.20
Syn. condenser	0.0	−30.00	27	0.00	0.00	0.0					
Syn. reactor	0.0	0.00	0			9.4	0.00	0.00			
Fault Pt. Amps									633	0.00	0.00

* Figure 6 is the circuit for which these values apply. The reference capacitor is connected to the bus G. For the fault operating condition, a 3-phase fault is applied to the bus E.

** Per cent amperes, 50-mva base.

† Deviation value expressed as percent amperes, 50-mva base. The degrees represent the angle between the current in the impedance-phase-shift network and the correct system value.

value of the incremental current $\boldsymbol{I_n}$ in a capacitor of $\boldsymbol{B_n}$ per-unit mhos, connected to a point in the network having a per-unit voltage to neutral of $\boldsymbol{V_{on}}$, is given by the equation

$$I_n = 0.0872(B_n)(V_{op} - V_{on})\underline{/177.5^\circ}, \quad \text{per-unit amperes} \quad (1)$$

where $\boldsymbol{V_{op}}$ represents the per-unit voltage to neutral of the network point to which the reference capacitor is connected. The corresponding per-unit incremental watts and vars in the capacitor circuit, W_n and $Vars_n$, due to this incremental current are given by the following equations:

$$W_n = 0.0872(B_n)(V_{on}^2)\times \left[1 + 1.001\,\frac{V_{op}}{V_{on}}\cos(177.5+\theta)\right], \quad \text{per-unit watts} \quad (2)$$

$$Vars_n = 0.0872(B_n)(V_{on}^2)\times \left[\frac{V_{op}}{V_{on}}\sin(177.5+\theta) - 0.0436\right], \quad \text{per unit vars} \quad (3)$$

where θ is the angle of the vector $\boldsymbol{V_{on}}$ subtracted from the angle of the vector $\boldsymbol{V_{op}}$, and the other symbols as defined for equation 1.

The incremental current and the incremental watts and vars of the capacitor $\boldsymbol{n}$, and also for any other capacitor, merge with the current, the watts and the vars passing the point at which the capacitor is connected. These incremental quantities are supplied by the various generating units, being divided among them in vector proportion directly as the vector admittances from the point of the capacitor attachment to the various points of internal voltage generation. The total incremental effect in any circuit element is the vector summation of the several incremental quantities in the element due to the several capacitors of the network.

Analyzer networks representing systems of 69- and 138-kv lines covering areas of 15,000 to 20,000 square miles will have cumulative incremental currents in the lines and the generators that will result in these currents deviating from their correct value by 0.1 per cent or less. The deviation in the vector angles of the currents will be 0.1 degree or less.

The deviation of the currents in the important parts of the 310-mile 161-kv system shown in Figure 6 for three operating conditions is tabulated in Table II. The reference capacitor for this circuit is connected to bus *G*. The 3-phase fault occurs on bus *E* during heavy-load conditions. For fault studies, the capacitors, in effect, are connected to the neutral bus since the capacitor generator's voltage is made zero during fault investigations.

Since the deviation values are much too small to be measured when in combination with the much greater line quantities, they were determined by calculating the various capacitor incremental currents by the use of equation 1 and then combined vectorially with the line currents of the various circuit elements. The differences in the magnitude of the element currents so obtained and the correct element currents for the three operating conditions are shown in columns 4, 7, and 10. The angles between the vectors representing these currents are shown in columns 5, 8, and 11.

The capacitor incremental currents during fault conditions are calculated by the equation

$$I_n = 0.0872(B_n)(V_{on}), \text{ per-unit amperes} \quad (4)$$

where $\boldsymbol{V_{on}}$ represents the voltage of the bus $\boldsymbol{n}$ during the faulted condition and the other symbols are those defined for equation one.

Summary

The impedance-phase-shift principle makes possible the use of reactors having *X-R* ratios of approximately eight at 60-cycle frequency which, however, function in the phase-shift network as though they possessed an *X-R* ratio of 25. By tolerating a slight increase in reactor weight, that corresponding to a 60-cycle reactor of 11.5 *X-R* ratio, they may be designed to have equivalent *X-R* ratios of infinity.

With the use of reactors having equivalent zero losses, it is no longer necessary to make a correction in the analyzer resistance value to compensate for the losses in the reactors being used. The resistance knobs and the reactance knobs are set to the identical values as those for the per cent resistance and reactance of the line impedance of the system network to be represented.

Also, because of the equivalent zero loss of the load-unit reactors a new type design for analyzer load units has become possible. This new design permits the direct setting up of the load in per cent kilowatts and per cent kilovars, thus obviating the necessity of determining the equivalent resistance and reactance required to represent a system load. A considerable saving of time is effected by this change in procedure.

Magnetic - type voltage-stabilizing transformers and static-type phase balancers replace the special-type motor-generator set, or the large number of electronic tubes and apparatus required for obtaining the power supply for higher than 60-cycle analyzers. The power for the stabilizing transformers is obtained directly from a 3-phase commercial supply line, no rotating equipment being required.

As a result of the negligible self-capacitance effect of the leads of analyzer circuits using 60-cycle frequency, a much higher ratio of base-volts to base amperes (base impedance) may be selected for the 60-cycle analyzer than may be used for analyzers operating on higher frequencies. For the same degree of self-capacitance effect, the ratio of the impedance base of a low frequency analyzer to the impedance base of a high-frequency analyzer may be made equal to the inverse ratio of their frequencies. The use of a relative impedance base of this value will result in equal size capacitors required in the analyzers to represent a given per cent system capacitance.

The degree of accuracy in the representation of capacitance effect of a system network on the impedance-phase-shift-type analyzer is several times greater than the accuracy of the standard instrument system, namely, for scalar values, 0.5 per cent of full-scale value of range in use, and for angle values, 0.5 degree. In other words, the deviation values indicated in Table II are of more interest from a theoretical viewpoint than from a practical one inasmuch as the differences in the indicated values and the correct values are not discernable on the instrument ranges required to read the system quantities.

Reference

1. IMPROVEMENTS IN SYSTEM ANALYZERS, Emory B. Phillips. United States Patent 2,487,942.

Discussion

R. B. Squires (Westinghouse Electric Corporation, East Pittsburgh, Pa.): This paper is an excellent exposition of how the phase shift principle is used to reduce the weight of reactors in network calculators. The reduction may not be as great as shown, since our design of 440-cycle generator reactors of the same range as shown in Table II of the paper have a *Q* of 30 to 50 and weigh only 10 pounds total instead of 53.6 pounds.

It must be realized, moreover, that the weight or size of components does not affect the total size of a calculator directly. The cost of an instrument system, of connection facilities, and of accurate calibration of all units, must be included.

In the design described in this paper, the interesting phase shift principle also requires that extra resistance be added for all taps of the reactors, and that all resistors have capacitors in parallel with them. These extra components must also be calibrated, thereby, increasing the over-all cost.

E. B. Phillips: Mr. Squires notes that the total weight of a 440-cycle group of reactors is only 10 pounds while those of the 500-cycle design, table I of the paper, total 53.6 pounds and, apparently, concludes from this that the weight reduction resulting from the use of the impedance-phase-shift principle may not be as great as shown. Squires should realize that he is making a comparison of reactor groups employing differing types of construction and that the reactor group that he cites as weighing 10 pounds does not have an average Q of 50, whereas the other group contains 12 reactors everyone of which has a Q of 50.

The weight of a reactor varies very rapidly with a change in its Q value, approximately as the third power of its Q value. Naturally enough the group that contains units that have Q's between 30 and 50 will weigh much less than the group containing units all of which have Q's of 50. Also, the 440-cycle design group employs the multisection-coil construction while the other group uses the single-section-coil construction that requires four reactors to give the same ohmic range and steps as a 10-section type reactor. The 10-section tope reactor, though costing more to wind, will weigh about one-half that of the equivalent group of four units.

In order to arrive at a somewhat correct weight comparison, suppose it be considered that the group of 12 reactors, each of which have a Q of 50 and shown in table I of the paper as weighing 2.08 pounds, are modified to the extent that they will have Q's varying between 30 and 50 and also modified to have the multisection-coil type construction. Their change in Q value will give them a group weight somewhere between the 1.56 pounds for the group having Q's of 25, see table I of the paper, and the 2.08 pounds, let us say 1.8 pounds. Then this weight must be halved, due to the change in the construction, which will result in a weight of 0.9 pound. Now it is this group of reactors that will have the same Q characteristics and the same type construction as the 440-cycle group. The weight reduction as a result of using the impedance-phase-shift principle is 10 pounds to 0.9 pound, a change in weight of 11 to 1.

With reference to the observations on the instrument system, accurate calibration of all units, et cetera, I think it holds without contradiction that all analyzers made by reputable and long-established manufacturers possess reliable and adequate accuracy in the calibration of all units, also reliable and accurate instrument systems. No one company enjoys an exclusive claim on this achievement. These items are clearly set out in the technical specifications and are not subject to conjecture. Incidentally and for your information, this analyzer is not one of those that was developed and built by an educational institution and thus having its total cost incorrectly determined as a result of nonrepresentative cost figures being used to cover the important items of development and labor costs.

In the matter of added costs of the extra resistors and capacitor units: since the reactors are not of the multisection winding type there are not some 30 to be added to the taps of the reactors, but instead, 11 are required for the line impedance units and 9 only for the load units. These resistors are of the form used in the final adjustment of instruments, a small fiber-type bobbin containing some 12 to 15 inches of resistance wire. No additional measurements are required to determine their values inasmuch as it always has been the customary practice to measure the equivalent resistance of all reactors upon the completion of their gap adjustment as a check on their Q value. This equivalent resistance is subtracted from a particular-value resistance that pertains for each reactor size, a resistance that is equal to the ohmic value of the reactor times the cotangent of 85 degrees.

For adjusting the resistors to a -5 degree angle, a single, stock-size type of ±10 per cent nominal value is required for only 6 of 11 line-unit resistors since the lower steps require a correction that is less than 0.1 per cent base vars. The adjustable, wire-wound tubular resistor, as regularly used for a resistance, serves as a potential divider for the capacitor. The simultaneous adjustment of the capacitor and the resistor requires but little more time than would be required for the resistor alone.

The cost of the trimmer resistors, the capacitors and the labor costs involved by them represents but 14 per cent of the total cost of a line impedance unit and but 10 per cent of the total cost of a load unit. Both of these costs are less than the savings resulting from the lower cost of the units in comparison to what their cost would be for 500-cycle designs. The principal cost reduction being due to the several pounds reduction in the costly core material and the remainder resulting from the reduction in the weight of copper required.

I am in full agreement with Squires in his remarks that a reduction in the weight of reactors can not result in a sizable reduction in the total cost of an analyzer, a reduction of only 6 to 8 per cent. However, with the use of the impedance-phase-shift principle, this reduction is not obtained at the expense of introducing some disadvantages, such as the requirement of complicated circuits and apparatus which has occurred in some instances wherein reactor weight reduction has been achieved.

Not only has the use of this new principle effected reactor weight reduction but its use has made possible some new and valuable features, such as:

1. A new type load unit that is calibrated in per cent kilowatts and kilovars which has reduced the time required to set the system loads on the analyzer.

2. A reduction in the time required for setting the system line impedances on the analyzer because of the zero-loss reactors.

3. A type load unit that permits of a more accurate setting on the analyzer of the important (larger value) system loads.

4. An improvement in the measurement system that eliminates the need for the hundreds of circuit selection relays (made possible only as the result of the low frequency base).

5. The elimination of all forms of frequency conversion apparatus. A compact, less costly power supply system, receiving its power directly from a 60-cycle supply line, is contained in the analyzer cabinet.

The Effect of Core Materials on Magnetic Amplifier Circuits

LEO J. JOHNSON
ASSOCIATE MEMBER AIEE

Synopsis: This paper shows the hysteresis loop characteristics exhibited by four different core materials while actually employed in a magnetic amplifier circuit. These characteristics serve to point up differences which affect the operation of magnetic amplifiers. It is hoped that from this information the designer will be able to apply these characteristics to better advantage as well as to increase the field of application.

A SERIES of new-type fast-response magnetic amplifier circuits were recently developed at the Naval Research Laboratory as a direct result of a new theoretical approach to the magnetic amplifier problem.[1,2] It is an example of this type of amplifier which is used to demonstrate the effects of core materials.

The operation of the new circuits depends upon the ability of the magnetic core material to remain at any given flux density level, up to saturation, in the absence of an applied magnetomotive force. Many core materials, which are generally considered superior for use in conventional magnetic amplifier circuits, do have the property of remaining at or near any existing flux level when applied magnetomotive forces are removed. However, as this paper demonstrates, one core material which is considered superior for use in magnetic amplifier circuits will not remain at an arbitrarily chosen flux level (saturation and below) in absence of an applied magnetomotive force. For this reason, this and similar core materials are theoretically inappropriate for use in the new type magnetic amplifier circuits and are shown experimentally to exhibit relatively poorer performance.

This paper demonstrates experimentally the magnetic characteristics which commercially produced core materials exhibit during their use in magnetic amplifiers. Experimental evidence is also used to show the effect of rectifier leakage upon the operation of the new-type circuits. The relative importance of rectifier leakage is observed when core materials having different loop widths are used and compared.

Choice of Circuit

The single core amplifier circuit shown schematically in Figure 1 is the circuit used for the series of experiments. The circuit includes a 2-winding transformer, a pair of rectifiers, a load, a control voltage source, and a 60-cycle a-c supply. The operation of this type of amplifier has been described in a previous publication[2] and therefore will not be repeated here.

In the case of the tests on the core materials reported herein, the magnetic path length, the number of turns, the size of wire, the rectifier current ratings, and load resistance were kept the same. A direct comparison of the four core materials is made for the given set of circuit parameters tabulated in Appendix I.

The cores are toroidal in shape and of wound tape construction. The tape width is one-half inch for core *A*, core *B*, and core *D*, and five-eighths of an inch for core *C*. The circuit of Figure 1 was chosen because of its simplicity and because it is the basic circuit of the series of new amplifier circuits. The basic operating principles of the circuit are the same as those of the more complex circuits which can be built with it.

The control voltage wave form is that of a continuous series of half sinusoids. The control voltage is obtained by rectifying the 60-cycle a-c supply through a variable voltage transformer by a full wave dry rectifier. By this means the control voltage wave form is made the same as that of the a-c supply to the control circuit and 180 degrees out of phase with it during the reset half-cycle.

Characteristics

General

The measured transfer characteristics for the circuit of Figure 1 using the parameters described in Appendix I are shown in Figure 2. The hysteresis loops for the various core materials are shown in Figures 4 through 11. The circuit used for the measurement of these loop characteristics is shown in Figure 3. The ratio of R_1/R_2 is the same as the turns ratio of the power winding to the control winding N_1/N_2. Thus the signal to the horizontal amplifier of the cathode-ray oscilloscope is proportional to the total magnetizing force (the integrator draws negligible current from the search coil). Figures 4 through 7 show the major hysteresis loops of the four core materials as taken by the circuit of Figure 3. The major loops were obtained by short circuiting the d-c control voltage terminals. Figures 8 through 11 show sets of superposed minor loops. The three minor loops shown are for control voltage settings of zero, an intermediate value, and a setting corresponding to a value just below the knee of the transfer characteristic for each material. For control voltage settings of 50 volts, maximum values shown in Figure 2, the minor loops disappear when using core *B*. The minor loops for core *D* exhibit practically no change in shape from the smallest of the three loops shown in Figure 11, and the minor loops for core *A* and core *C* get quite small but do not disappear when a control voltage of 50 volts is applied. It may be noted that in all four cases, a control voltage setting of 50 volts is equivalent to open

---

Paper 52-26, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, January 21–25, 1952. Manuscript submitted October 24, 1951; made available for printing December 31, 1951.

Leo J. Johnson is with the Naval Research Laboratory, Washington, D. C.

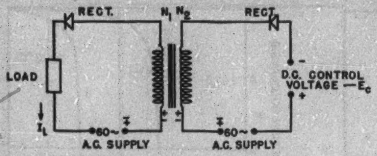


Figure 1 (left). Schematic diagram of a single core magnetic amplifier

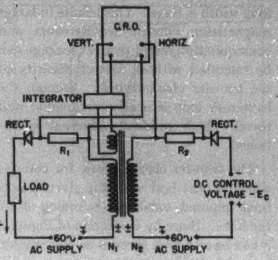


Figure 3. Schematic diagram showing means of measuring hysteresis loops

circuiting the control circuit, since with this setting the net voltage in the control circuit is always blocked by the control circuit rectifier.[2]

A pronounced bulging and sharpening of the second quadrant knee is shown for the minor loops of core *A*, core *B*, and core *C* while this effect is negligible for core *D*. When this phenomenon was noted, special precautions were applied to make sure that the measuring techniques did not affect the result. It was found that the major part of this bulging is a characteristic of the core when operating in the magnetic amplifier circuit used in the experiments. As can be seen in Figures 8 through 11, the cores exhibit dynamic characteristics quite different from those which are obtained with small constant values of d-c bias magnetomotive force and reduced a-c applied voltage.[3] This is probably due to the fact that the peak values of magnetomotive force are larger by a factor of more than 1,000 in actual amplifier operation.

Correlation of Loop Characteristics to Transfer Characteristics

A scrutiny of the transfer characteristics shown in Figure 2 reveals discrepancies from the predicted ideal amplifier of the theory.[1] There are three major reasons for the experimental transfer characteristic to deviate from the theoretical transfer characteristic. First matching of a-c supply voltages in the power and control circuits is quite critical. A slight error in the ratio of a-c supply voltage of the power circuit to that of the control circuit will shift the transfer characteristic. If the a-c supply voltage in the control circuit is too large, the shift will be to the left and if the a-c supply voltage is not large enough, the shift will be to the right. Second, operation of the new circuits depends upon the maintenance of flux at saturation level without need of an applied magnetomotive force. If this condition is not met, the amplifier will reset without regard to the control voltage.[1] This reset will prevent the amplifier from having complete half-cycle conduction and consequently the power handling capacity will be reduced. Third, but not least, is rectifier leakage. Leakage through the load circuit rectifier will produce a magnetomotive force which aids the magnetomotive force derived from the control circuit current. Therefore, if, and only if, the rectifier leakage (times the turns ratio) tends to exceed the control circuit current requirements for core magnetization, reset will be accomplished without regard to the control voltage. This results in a reduction of power handling capacity and loss of linearity of the transfer characteristic. With the use of improved rectifiers, the effect of leakage can be minimized and core materials with narrower loops can be used with a consequent higher gain factor and without loss of power handling capacity or transfer characteristic linearity.

An ideal transfer characteristic is assumed to be one which exhibits a linear relationship of average load current versus average control volts with the load current ranging from zero to some maximum value called saturation. Ideally, independence of supply frequency and voltage, rectifier leakage, and core hysteresis loop width is implied.[1] This is considerably more independence than can generally be obtained in conventional circuits where control current is used as the independent variable. It is assumed that the control voltage is derived from a voltage similar in wave shape and of the same phase as the a-c supply voltage to the control circuit.[2]

The transfer characteristic for core *B* most nearly approaches the ideal for the following reasons. First, the hysteresis

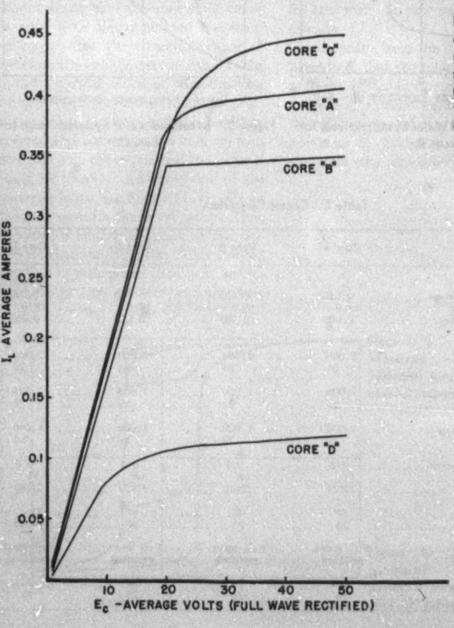


Figure 2. Transfer characteristics of single core amplifier showing load current (I_L) versus control voltage (E_c) for four different core materials

loop width is large. This results in larger magnetizing current requirements and consequently greater rectifier leakage can be tolerated without loss of linearity of the transfer characteristic. Second, the hysteresis loop is nearly rectangular and thereby high residual induction is obtained.

The transfer characteristic for core *A*, which is next best with the given conditions imposed, exhibits a rounding off at the knee. Since by inspection of Figure 8 it can be seen that there is no appreciable loss of induction at remanence, it is deduced that the load circuit rectifier leakage is greater than half the width of the hysteresis loop. This deduction is borne out by the oscillographic records shown in Figure 12. The top oscillogram is an indication of the load circuit rectifier leakage while the curve just below it is an oscillogram of the control circuit current, both taken when the control voltage was set at 20 volts. The sharp peaks of current at the extreme right of the oscillograms are to be disregarded at this time. A reference to the transfer characteristic in Figure 2 shows 20 volts to be the point where nonlinearity begins. A reference to Figure 12 shows that the rectifier leakage current (top trace) is just equal to the control current requirement at one point. In other words, the peak rectifier leakage is about to interfere with the controlled reset. The third oscillogram of Figure 12 shows the control current trace when the load circuit rectifier leakage current is zero for comparison purposes. Thus it can be seen that since core *A* has a narrow hysteresis loop the use of a rectifier with less leakage in the load circuit would markedly improve the transfer characteristic.

The transfer characteristic for core *C*, third best of the group, shows an even more pronounced rounding at and above the knee. The rounding effect can be caused by one or both of two things. Loss of induction at remanence and rectifier leakage. An inspection of the minor loops of Figure 10 shows only a very small loss of induction. In fact, it is to be noted that remanence induction is slightly higher for the minor loops. It has been suggested that since the core material is relatively thick for this sample, eddy currents may have a prominent effect on the loops. The same rectifier was used in both cases and rectifier leakage increases with voltage. Therefore, it is concluded that rectifier leakage is the major contributing factor since a supply voltage greater than that used for core *A* was employed. This conclusion is borne out by the two transfer characteristics shown in

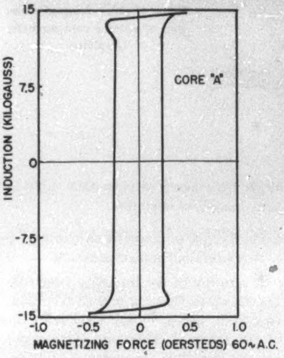


Figure 4. Measured major hysteresis loop for core A

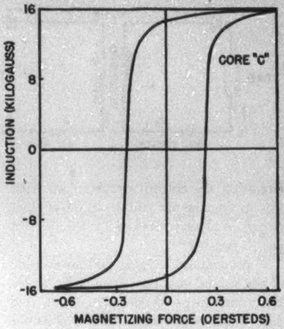


Figure 6. Measured major hysteresis loop for core C

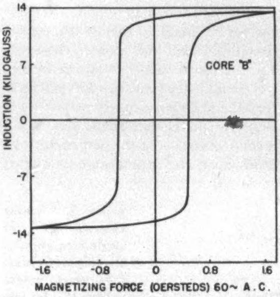


Figure 5. Measured major hysteresis loop for core B

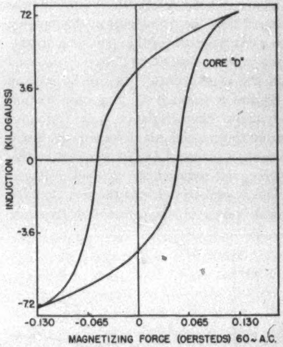


Figure 7. Measured major hysteresis loop for core D

Table I. Circuit Parameters

Parameter	Core A	Core B	Core C	Core D
Core				
Area, square centimeters	0.720	0.658	0.720	0.733
Length, centimeters	18.0	18.0	18.0	18.0
Strip thickness, inches	0.002	0.002	0.012	0.002
Coils				
Power winding				
Turns	2,000	2,000	2,000	2,000
Size	24	24	24	24
Control winding				
Turns	1,000	1,000	1,000	1,000
Size	29	29	29	29
Search coil				
Turns	1,000	1,000	1,000	1,000
Size	29	29	29	29
Power rectifiers				
Current rating, amperes	0.5	0.5	0.5	0.5
Voltage rating	68	68	68	34
Control rectifier	IN39	IN39	IN39	IN39
A-c supply voltage				
Power circuit	57	47	67	28
Control circuit	29	24	34	14.25
Type of control voltage	Full wave rectified	Full wave rectified	Full wave rectified	Full wave rectified
Resistance, ohms				
Load	25	25	25	25
Control source	75	75	75	75

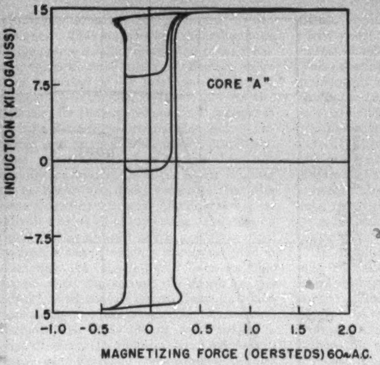


Figure 8 (left). Measured minor hysteresis loops for core A superposed

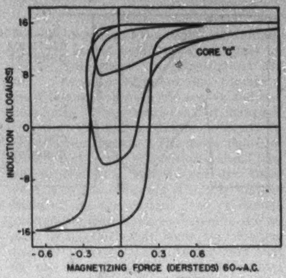


Figure 10 (right). Measured minor hysteresis loops for core C superposed

Figure 13. The two characteristics were obtained using core *C*. One is a replot of the transfer characteristic shown in Figure 2 while the other, which approaches the ideal characteristic, was obtained when a rectifier with less leakage current was used. In this latter case the rectifier leakage never reached the current value necessary for magnetization.

The transfer characteristic for core *D* deviates drastically from an ideal characteristic. Ideally, all four transfer characteristics would have the same slope below the knee since the same load resistance was used. Also, the saturation values of load current would be in proportion to the applied (supply) voltages. In the case of core *D* the slope of the transfer characteristic below the knee is much less than expected and the saturation value of load current is only about one-half the value it should be ideally. An inspection of the hysteresis loops, see Figure 11, reveals an obvious reason for the discrepancy. This, of course, is loss of induction at zero applied magnetomotive force. This marked deviation of remanence from peak induction affects the transfer characteristic throughout its range. An important secondary reason is rectifier leakage. Even a slight amount of rectifier leakage will reduce the induction further since the loop is quite narrow and slopes down quite markedly throughout the second quadrant and a low value of magnetomotive force due to rectifier leakage would accomplish uncontrolled reset. A considerable loss of power handling capacity results and for these reasons core *D* is considered an inappropriate core material for use with the new magnetic amplifier circuits.

It appears, from the experimental data presented, that the relatively poorer core material (wider hysteresis loop) obtains a more ideal transfer characteristic. It may be well to note that control current is proportional to hysteresis loop width and therefore the amplifier which employs cores with narrower hysteresis loops will have an inherently higher gain factor.[1] To maintain a nearly ideal transfer characteristic while employing narrower loop core materials, it is necessary to use rectifiers with less leakage, see Figure 13.

Concluding Remarks

The comparisons of the four core materials are based on experiment. The parameters contained in the circuit were not theoretically ideal. The major nonideal element, other than the cores, is the load circuit rectifier. Since theoretically ideal rectifiers have not as yet been produced, it is necessary to include the effects of nonideal rectifiers to be realistic. Fortunately, with proper choice of rectifiers, commensurate with the core materials used, an ideal transfer characteristic can be approached.

Since the control circuit current is equal to the magnetizing current minus the load circuit rectifier leakage current times the turns ratio, the criteria for

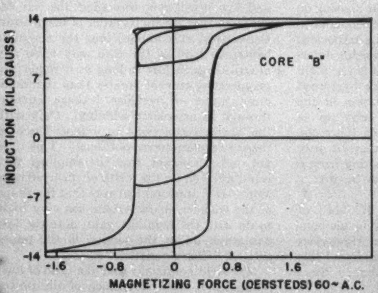


Figure 9 (left). Measured minor hysteresis loops for core B superposed

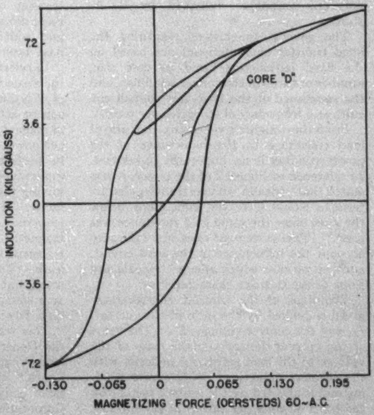


Figure 11 (right). Measured minor hysteresis loops for core D superposed

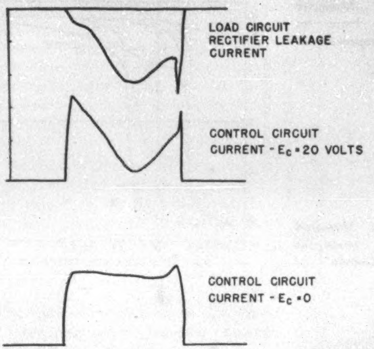


Figure 12 (left). Oscillographic traces of rectifier leakage current and control circuit current using core A

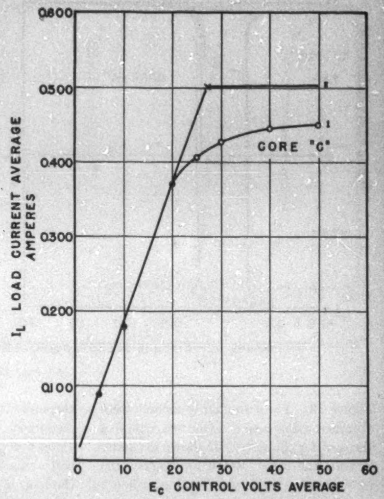


Figure 13 (right). Transfer characteristics obtained using core C

1. Replot of characteristic shown in Figure 2
2. Characteristic obtained when using a load circuit rectifier having less leakage

selecting an appropriate rectifier to obtain an optimum transfer characteristic is only that the leakage must not exceed the magnetizing current requirement, see Figures 2, 12, and 13.

It is recognized that the circuit used is new and unconventional. However, it is believed that this basic circuit, see Figure 1, with its inherent 1 cycle response and high gain factor is sufficiently important to be used for this investigation. The results of this investigation can be used in the general field of magnetic amplifier design and analysis. It is well to note, however, that the specific points of comparison made in this paper apply directly to the circuit, and its operation, given in Figure 1.

References

1. On the Mechanics of Magnetic Amplifier Operation, **Robert A. Ramey.** *AIEE Transactions*, volume 70, part II, 1951, pages 1214–23.

2. On the Control of Magnetic Amplifiers, **Robert A. Ramey.** *AIEE Transactions*, volume 70 part II, 1951, pages 2124–28.

3. Properties of Deltamax, 4-79 Mo-Permalloy, Supermalloy, *Bulletin TC-100.* The Arnold Engineering Company (Marengo, Ill.), March 15, 1951.

Discussion

William J. Field (Engineering Research Associates, Inc., St. Paul, Minn.): This paper gives a very useful insight into the operational effects of core magnetization and rectifier leakage currents on the types of magnetic amplifiers described by R. A. Ramey.

The general conclusions regarding the ideal transfer characteristics are based on the fixed parameters listed as core size, number of turns in the copper windings, and the resistance of the load and control circuits at a frequency of 60 cycles per second.

From the designer's viewpoint, the ratio of load resistance to the inductance of the power winding is an important parameter. In reference to Figure 2 of the paper, it was stated that "Ideally, all four transfer characteristics would have the same slope below the knee since the same load resistance was used." This statement does not take into account the inductance in the load circuit, another variable which affects the scale and slope of the transfer characteristic

The slope of the transfer characteristic given is defined by the ratio of the current, I_L, and the control voltage, E_C. The value of the current depends on the ratio of the voltage to the load resistance in series with the effective inductance of the power winding and is affected by the control voltage on the basis of volts per turn. If the number of turns in the copper windings on core *D* were decreased, the current I_L would have to increase and control voltage E_c and would have to decrease proportionately to maintain the same magnetization of the core. I believe that if the flux linkages per ampere with core *D* had been more nearly comparable to those obtained with core *B*, the transfer characteristics would have been more alike and the effects of the squareness and width of the hysteresis loops would have been more representative.

In reference to the pronounced bulging at the second quadrant knee as shown for core *A*, it is difficult to understand that the cause can be attributed to the larger peak values of magnetomotive forces. If the phasing on the power and control sides is considered to be perfect, it may be that the unnoticed inductance present in the resistor for determining the magnetizing current is the main factor. The squareness of the hysteresis loop indicates an abrupt deviation in the magnetizing current. This sets up a transient in the resistor between the scope leads. The stray inductance may be very small, but the components resulting from a very square loop at this point can be of very high frequencies.

This was indicated by Mr. H. Lord of the General Electric Company in his conference paper on "Dynamic Hysteresis Loops." Mr. D. C. Dieterly has found this same cause in his work at Armco Steel Corporation.

Leo J. Johnson: The discussion presented by Mr. Field is welcomed by the author since it gives him an opportunity to expand and clarify some possibly obscure points.

First, it seems necessary to point out that the general conclusions as to the ideal transfer characteristics are not based upon specific circuit parameters. The general conclusions should be based more on the compatibility of any set of parameters. An example of this can be drawn from the second paragraph under the section in the paper entitled "Concluding Remarks." This paragraph means that the rectifier leakage current may be any value whatsoever as long as it is less than that value required for core magnetization and if so, it will in no way effect the transfer characteristic within its order of magnitude. Conversely, this implies that the magnetic hysteresis loop of the core may have any practical width just so long as it requires a magnetizing current greater than the maximum value of rectifier leakage current through its associated winding. Only with the above restrictions is independence of these two parameters obtained. This does not mean, however, that the amplifier gain is not affected by the width of the hysteresis loop. It is important to note that the shape of the transfer characteristic has very little to do with the amplifier gain, note the last paragraph under the section in the paper entitled "Characteristics."

Mr. Field points out that the ratio of load resistance to the inductance of the power winding is an important parameter to the designer. This is true when the inductance of the power winding is assumed to be linear and equal to some mean or effective value over a period of at least one cycle. It would be well to note, however, that the circuit used for the experiments in this paper is a result of a new theoretical approach to the design and analysis of magnetic amplifiers. This new theoretical approach does not assume the inductance of the power winding to be linear nor does it assume that the power winding has any specific average or effective value of inductance. The circuits developed as a result of the new theoretical approach, one of which is the circuit used in obtaining the experimental data for this paper (note the second paragraph of the paper), must contain circuit elements which reasonably approximate the theoretical assumptions. One of the assumptions is that the magnetic material has a very small value of differential permeability beyond the knee of the magnetic saturation characteristic. In the case of core D space did not permit a complete representation of the saturation characteristic. However, in clarification of this the author wishes to point out that the differential permeability in the region of magnetomotive force equal to 0.3 oersted for core D is of the same order of magnitude as for the other core samples used. Thus, the assumption of small differential permeability beyond the knee of the saturation characteristic is adequately met. However, one of the theoretical assumptions is being unreasonably violated, namely, the assumption that the value of flux density must not change when an applied magnetomotive force is removed. The inherent properties of the material of core D are such that the flux density necessarily must change appreciably upon removal of an applied magnetomotive force. This inherent characteristic is the one which renders the core material inappropriate for use in the new type amplifier circuits. The loss of induction is the reason that the transfer characteristic obtained when using core D, has less slope below the knee than the other transfer characteristics shown in Figure 2 of the paper. Loss of induction is not equivalent to having inductance in the load circuit, even though, such inductance would have a similar effect on the average values.

In regard to Mr. Field's statement about flux linkages per ampere (last sentence of third paragraph in his discussion) the author wishes to point out that when the number of turns are reduced the a-c supply voltages must be reduced proportionately. When this is done, the current I_1 will not increase for a given value of E_c. The net result will be that the transfer characteristic will be the same below the knee and the knee will occur at a lower value of I_1 and E_c than it occurred originally. The simple experiment of changing the number of turns bears this out. From the above it can be surely determined that the value of load resistance along with the turns ratio determines the slope of the transfer characteristic below the knee provided the basic assumptions are met within practical limits.

In reply to the last paragraph of Mr. Field's discussion it may be pointed out that special precautions were applied to make sure that the measuring techniques did not effect the result (see the second paragraph of the section in the paper under "Characteristics"). The precautions included the checking of linearity, saturation, and transient response of the oscilloscope amplifiers, the use of noninductive resistors as well as close coupled windings about the core. The resistor in question contained about two microhenrys and twelve and one half ohms. The measured rate of change of current at the time in question is nearly 36 amperes per second. The peak value of current at the point in question is about four milliamperes (the values used are those obtained when using core A). The voltage produced by the current rate of change through the inductance contained in the resistor is 0.000072 volt while the resistive drop is 0.05 volt. These values show that the voltage contribution of the inductance contained in the measuring resistor is less, by three orders of magnitude, than the resistive voltage contribution and therefore can be safely neglected. Subsequent to the Winter General Meeting in New York, the author has received a copy of Mr. H. Lord's conference paper on "Dynamic Hysteresis Loops" and a letter from Mr. H. Lord. In his letter Mr. Lord states:

"We have observed similar bulging at this outer knee. Certain materials show it much more pronounced than others when viewed in the same test circuit, so I believe it is truly a magnetic material phenomena."

The author has shown in the paper several oscillographic representations of symmetrical and assymetrical hysteresis loops Three of the four core materials show the bulging at the knee in different degrees of shape and magnitude. From these facts, it is quite reasonable to believe that this phenomenon of bulging at the second quadrant knee is a characteristic of the magnetic material.

Electromechanical Analogies of a Separately Excited D-C Machine

R. E. VOWELS
MEMBER AIEE

W. G. FORTE
NONMEMBER AIEE

Synopsis: The method of electromechanical analogies is now well established, the three types being the direct analogy, the inverse analogy, and the frequency analogy. This paper reviews the first two methods and applies them to the transient analysis of the separately excited d-c machine. The equivalent circuits developed allow for motion of the stator and permit the analysis of vibration problems. Reference is made to the introduction of initial conditions.

THE object of electromechanical analogies is the establishment of a correspondence between a mechanical system and an electric system, element for element and system for system. Where lumped elements can be used, the electric circuit analogue is attractive because of the well-developed state of electric circuit theory with its almost universal notation, and the comparative ease in constructing, modifying, and taking measurements on an electric circuit model.

In one method, the direct analogy, a correspondence is made between electromotive forces and mechanical forces (or torques) and between currents and velocities. The other method, the inverse analogy, makes correspondence between electromotive forces and velocities and also between currents and mechanical forces.

In both these analogies, the frequencies of oscillation of the equivalent electric system are the same as those of the mechanical system. The equality of frequencies between the two systems is not an essential requirement, and Cahen[1] has shown that it is possible to establish a double infinity of electric network analogies of a given mechanical system by making the frequency of mechanical oscillations correspond to a function of the frequency of oscillation of the electric system. This might be referred to as the frequency analogy.

When using analogies to determine the transient response of a system, it is a simple matter to introduce the initial conditions; for example, masses with initial velocities and springs with initial tensions may be represented by charges on capacitors or currents in inductances.

As a general rule, it is advantageous to use mechanical models at low frequencies where loss-free components are more readily obtainable, whereas at high frequencies electrical models are to be preferred. Nonlinear effects can be introduced into electric circuits and permit the study of nonlinear problems which are difficult or tedious to solve by computation. There should be much scope in the future for electrical analogies using these nonlinear effects.

The purpose of this paper is to present the equivalent circuit or electrical analogue of a particular electromagnetic transducer, namely, the shunt d-c machine. The appropriate analogy is the inverse, but once the equivalent electric circuit is established, it is a simple matter to convert to the dual which is equivalent to the use of the direct analogy.

Special care must be taken in drawing the duals of nonplanar networks, but reference may be made to a paper by A. Bloch.[2]

Initial Conditions

To demonstrate the way in which the initial conditions may be introduced, the principles of the direct and inverse analogies are re-stated[3-8] and the equations written in the Laplace transform notation. The series and parallel circuits containing resistance, inductance, and capacitance in the electric systems, and viscous resistance, mass, and compliance in the mechanical systems, are shown in Figure 1.

The voltage equation for the resistance R, inductance L, capacitance C, series circuit of Figure 1(A) set up with the aid of Kirchhoff's Second Law is

$$E=E_R+E_L+E_C=RI+L\frac{dI}{dt}+\frac{1}{C}\int I\,dt \tag{1}$$

where

E denotes electromotive force
I denotes current
t denotes time

At time $t=0$, the switch is closed with the initial conditions that the charge Q_t on the capacitor C is Q_{t0} and the current $I=I_0$.

Replacing each term of equation 1 by its Laplace transform gives the subsidiary equation:

$$\mathbf{E}=\left(R+Lp+\frac{1}{Cp}\right)\mathbf{I}-LI_0+\frac{Q_{t0}}{Cp} \tag{2}$$

where

$$\mathbf{E}=\int_0^\infty e^{-pt}E\,dt$$

and

$$\mathbf{I}=\int_0^\infty e^{-pt}I\,dt$$

p being a complex number such that the $Re[p]>k>0$, where k is a real number sufficiently large to make the infinite integrals convergent.

The corresponding current equations for the R', C', L' parallel circuit of Figure 1(B) set up with the aid of Kirchhoff's

---

Paper 52-45, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted June 14, 1951; made available for printing January 2, 1952.

R. E. Vowels and W. G. Forte are both with the University of Adelaide, Adelaide, South Australia.

The authors wish to acknowledge the initiative and encouragement of Professor E. O. Willoughby, Head of the Electrical Engineering Department, who established the Servomechanisms Laboratory in which this work was carried out.

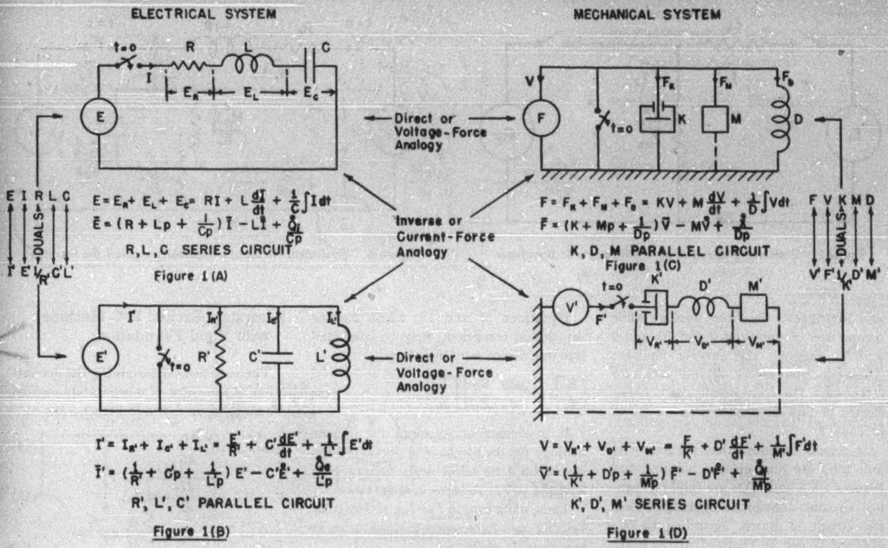

ELECTRICAL SYSTEM
MECHANICAL SYSTEM
t=o
R
L
C
I
E_R
E_L
E_C
E
Direct or Voltage-Force Analogy
Inverse or Current-Force Analogy
V
F
F_K
F_M
F_D
K
M
D
E I R L C
DUALS
I' E' 1/R' C' L'
F V K M D
V' F' 1/K' D' M'
$E = E_R + E_L + E_C = RI + L\frac{dI}{dt} + \frac{1}{C}\int I\,dt$
$\bar{E} = (R + Lp + \frac{1}{Cp})\bar{I} - L\overset{\circ}{I} + \frac{\overset{\circ}{Q_i}}{Cp}$
R,L,C SERIES CIRCUIT
Figure 1(A)
$F = F_K + F_M + F_D = KV + M\frac{dV}{dt} + \frac{1}{D}\int V\,dt$
$\bar{F} = (K + Mp + \frac{1}{Dp})\bar{V} - M\overset{\circ}{V} + \frac{\overset{\circ}{x}}{Dp}$
K, D, M PARALLEL CIRCUIT
Figure 1(C)
I'
$I_{R'}$
$I_{C'}$
$I_{L'}$
E'
R'
C'
L'
V'
F'
K'
D'
M'
$V_{K'}$
$V_{D'}$
$V_{M'}$
$I' = I_{R'} + I_{C'} + I_{L'} = \frac{E'}{R'} + C'\frac{dE'}{dt} + \frac{1}{L'}\int E'dt$
$\bar{I}' = (\frac{1}{R'} + C'p + \frac{1}{L'p})E' - C'\overset{\circ}{E}' + \frac{\overset{\circ}{Q_e}}{L'p}$
R', L', C' PARALLEL CIRCUIT
Figure 1(B)
$V = V_{K'} + V_{D'} + V_{M'} = \frac{F}{K'} + D'\frac{dF'}{dt} + \frac{1}{M'}\int F'dt$
$\bar{V}' = (\frac{1}{K'} + D'p + \frac{1}{Mp})F' - D'\overset{\circ}{F}' + \frac{\overset{\circ}{Q_f}}{Mp}$
K', D', M' SERIES CIRCUIT
Figure 1(D)

Figure 1. Direct and inverse electromechanical analogies

First Law are

$$I' = I_{R'} + I_{C'} + I_{L'} = \frac{E'}{R'} + C'\frac{dE'}{dt} + \frac{1}{L'}\int E'\,dt \qquad (3)$$

and

$$\mathbf{I}' = \left(\frac{1}{R'} + C'p + \frac{1}{L'p}\right)\mathbf{E}' - Q_0 + \frac{I_{L'0}}{p} \qquad (4)$$

where

$Q_0 = C'E'_0$ is the initial charge on the capacitor due to applied potential E'_0

and

$I_{L'0} = \left[\frac{1}{L'}\int E'dt\right]_{t=0} = \frac{Q_{e0}}{L'}$ = current in the inductance L at time $t=0$

Thus $\mathbf{I}'$ may be written in the form

$$\mathbf{I}' = \left(\frac{1}{R'} + C'p + \frac{1}{L'p}\right)\mathbf{E}' - C'E'_0 + \frac{Q_{e0}}{L'p} \qquad (5)$$

The equations for a viscous resistance K, compliance D, mass M, "parallel" type mechanical system shown in Figure 1(C) are

$$F = F_K + F_M + F_D = KV + M\frac{dV}{dt} + \frac{1}{D}\int V\,dt \qquad (6)$$

where

F denotes force
V denotes velocity

and

$$\mathbf{F} = \left(K + Mp + \frac{1}{Dp}\right)\mathbf{V} - MV_0 + \frac{x_0}{Dp} \qquad (7)$$

where

V_0 is the velocity of mass M at time $t=0$
x_0 is the displacement of spring D from the equilibrium position

The equations for a K', D', M' "series" type mechanical system shown in Figure 1(D) are

$$V' = V_{K'} + V_{D'} + V_{M'} = \frac{F'}{K'} + D'\frac{dF'}{dt} + \frac{1}{M'}\int F'dt \qquad (8)$$

and

$$\mathbf{V}' = \left(\frac{1}{K'} + D'p + \frac{1}{M'p}\right)\mathbf{F}' - Q_0 + \frac{V_{M'0}}{p} \qquad (9)$$

where

$Q_0 = D'F'_0$ might be termed the "charge" of the spring at $t=0$ due to force F'

and

$V_{M'0} = \left[\frac{1}{M'}\int F'dt\right]_{t=0} = \frac{Q_{f0}}{M'}$ = velocity of the mass M at time $t=0$

Thus $\mathbf{V}'$ may be written in the form:

$$\mathbf{V}' = \left(\frac{1}{K'} + D'p + \frac{1}{M'p}\right)\mathbf{F}' - D'F'_0 + \frac{Q_{f0}}{M'p} \qquad (10)$$

Comparison of equations 2 and 7, and 5

Table I. Correspondences between Electrical and Mechanical Systems

Mechanical System	Analogous Electrical Quantity	
	Direct Analogy	Inverse Analogy
Force F	Electromotive force E	Current I
Velocity V	Current I	Electromotive force E
Displacement $x = \int V\,dt$	Charge $Q_i = \int I\,dt$	$Q_e = \int E\,dt$
Impulse $Q_f = \int F\,dt$	$Q_e = \int E\,dt$	Charge $Q_i = \int I\,dt$
Viscous resistance K	Resistance R	Conductance $1/R$
Viscous conductance $1/K$	Conductance $1/R$	Resistance R
Mass M	Inductance L	Capacitance C
Compliance D	Capacitance C	Inductance L
Initial velocity of mass M	Initial current in inductance L	Initial charge on capacitor C
Initial compression or tension of spring (compliance) D	Initial charge on capacitor C	Initial current in inductance L

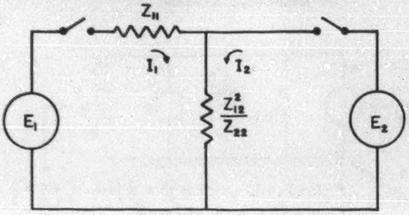


Figure 2. Elementary circuit of an electromagnetic transducer

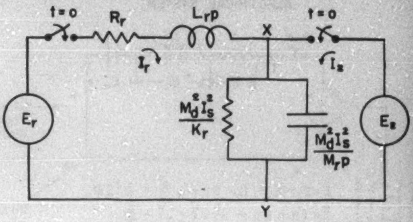


Figure 4. Equivalent circuit of a separately excited d-c machine

and 10, suggests the direct analogy while comparison of equations 2 and 10, and 5 and 7, suggests the inverse analogy. These equations are set out in Figure 1 for reference. The correspondences between electric and mechanical circuits are shown in Table I.

It should be noted that Figures 1(A) and 1(B) are dual electric circuits, and Figures 1(C) and 1(D) are dual mechanical circuits. Once the equivalent electric circuit is drawn, according to one analogy, it can be changed to the other analogy by drawing the dual of the first circuit.

Electromechanical Transducer

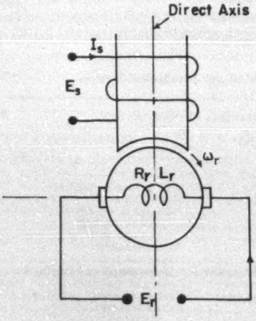
Assuming zero initial conditions, it may be shown that the equations governing the behavior of an electromechanical transducer utilizing electromagnetic forces take the subsidiary form[4]

$$\mathbf{E}_1 = Z_{11}\mathbf{I}_1 + Z_{12}\boldsymbol{\omega}_2 \text{ (voltage equation)} \quad (11)$$

$$\mathbf{T}_2 = -Z_{12}\mathbf{I}_1 + Z_{22}\boldsymbol{\omega}_2 \text{ (torque equation)} \quad (12)$$

where

E = electromotive force
T = torque
I = current
ω = angular velocity
Z = impedance expressed as a function of p



Equations 11 and 12, which assume zero initial conditions, may be combined into one matrix equation, thus

$$\begin{bmatrix}\mathbf{E}_1\\ \mathbf{T}_2\end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12}\\ -Z_{12} & Z_{22}\end{bmatrix}\begin{bmatrix}\mathbf{I}_1\\ \boldsymbol{\omega}_2\end{bmatrix} \quad (13)$$

It is convenient to adopt the inverse analogy for the mechanical portion of the system and to select scale factors such that the power remains invariant.

Thus, write torque $T_2 = I_2 a$, and angular velocity $\omega_2 = E_2/a$, and on substitution in equation 13, we have

$$\begin{bmatrix}\mathbf{E}_1\\ \mathbf{I}_2 a\end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12}\\ -Z_{12} & Z_{22}\end{bmatrix}\begin{bmatrix}\mathbf{I}_1\\ \mathbf{E}_2/a\end{bmatrix} \quad (14)$$

Equation 14 may be written in the form

$$\begin{bmatrix}\mathbf{E}_1\\ \\ \mathbf{E}_2\end{bmatrix} = \begin{bmatrix} Z_{11}+\dfrac{Z_{12}^2}{Z_{22}} & \dfrac{aZ_{12}}{Z_{22}}\\ \dfrac{aZ_{12}}{Z_{22}} & \dfrac{a^2}{Z_{22}}\end{bmatrix}\begin{bmatrix}\mathbf{I}_1\\ \\ \mathbf{I}_2\end{bmatrix} \quad (15)$$

If a is selected equal to Z_{12}, then

$$\begin{bmatrix}\mathbf{E}_1\\ \\ \mathbf{E}_2\end{bmatrix} = \begin{bmatrix} Z_{11}+\dfrac{Z_{12}^2}{Z_{22}} & \dfrac{Z_{12}^2}{Z_{22}}\\ \dfrac{Z_{12}^2}{Z_{22}} & \dfrac{Z_{12}^2}{Z_{22}}\end{bmatrix}\begin{bmatrix}\mathbf{I}_1\\ \\ \mathbf{I}_2\end{bmatrix} \quad (16)$$

which equation can be represented by an electric circuit of the simplest form shown in Figure 2.

Figure 3 (left). Separately excited d-c machine

Figure 5 (below). Flexible connection of load to rotor

Separately Excited D-C Machine with Rigid Foundation

The equations determining the acceleration of the rotor of a separately excited d-c machine as shown in Figure 3 are

$$E_s = R_s I_s + L_s \frac{dI_s}{dt} \quad (17)$$

$$E_r = R_r I_r + L_r \frac{dI_r}{dt} + M_d I_s \omega_r \quad (18)$$

$$T_r = -M_d I_s I_r + K_r \omega_r + M_r \frac{d\omega_r}{dt} \quad (19)$$

where

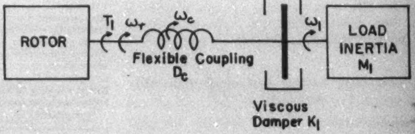
E_s, E_r are the stator and rotor electromotive forces respectively
I_s, I_r are the stator and rotor currents respectively
R_s, R_r are the stator and rotor resistances respectively
L_s, L_r are the stator and rotor inductances respectively
K_r, M_r are the viscous resistance and moment of inertia of the rotor, respectively
T_r, ω_r are the torque and angular velocity of the rotor, respectively
M_d is the mutual inductance of stator and armature circuits when stator and brush axes are coincident

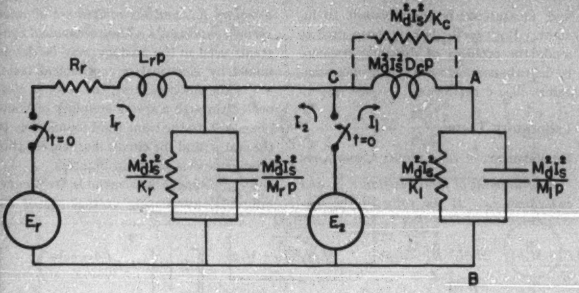
If the field winding is excited from a separate d-c source such that I_s is constant, then the subsidiary matrix equation is

$$\begin{bmatrix}\mathbf{E}_r\\ \mathbf{T}_r\end{bmatrix} = \begin{bmatrix} R_r + L_r p & M_d I_s\\ -M_d I_s & K_r + M_r p\end{bmatrix} \times \begin{bmatrix}\mathbf{I}_r\\ \boldsymbol{\omega}_r\end{bmatrix} - \begin{bmatrix} L_r I_{r0}\\ M_r \omega_{r0}\end{bmatrix} \quad (20)$$

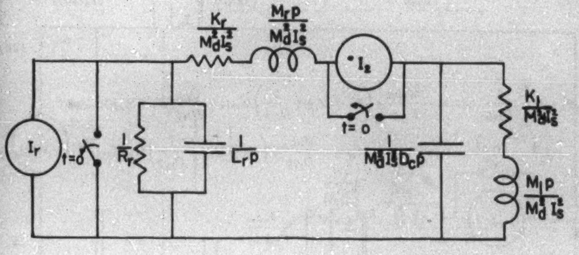
where

$I_r = I_{r0}$ at $t = 0$
$\omega_r = \omega_{r0}$ at $t = 0$





Original
(A)



Dual
(B)

Figure 6. Equivalent circuit of a separately excited d-c machine with flexibly connected load

Putting $T_r=I_2a$, $\omega_r=E_2/a$, and selecting $a=M_dI_s$ such that power remains invariant, then equation 20 may be written in the form

$$\begin{bmatrix}\mathbf{E}_r\\ \mathbf{E}_2\end{bmatrix}=\begin{bmatrix}R_r+L_rp+\dfrac{M_d^2I_s^2}{K_r+M_rp} & \dfrac{M_d^2I_s^2}{K_r+M_rp}\\ \dfrac{M_d^2I_s^2}{K_r+M_rp} & \dfrac{M_d^2I_s^2}{K_r+M_rp}\end{bmatrix}\times\begin{bmatrix}\mathbf{I}_r\\ \mathbf{I}_2\end{bmatrix}+\begin{bmatrix}\dfrac{M_rE_{20}}{K_r+M_rp}-L_rI_{r0}\\ \dfrac{M_rE_{20}}{K_r+M_rp}\end{bmatrix} \quad (21)$$

The equivalent circuit which represents matrix equation 21 precisely is shown in Figure 4.

It should be noted that:

1. The resistance and inductance of the rotor winding as measured with locked rotor are represented by the lumped circuit constants R_r and L_r, respectively.

2. Because of the use of the inverse analogy the moment of intertia of the rotor circuit M_r is represented by capacitance $1/M_rp$ multiplied by the square of the flux linkages M_dI_s.

3. The viscous damping of the rotor circuit K_r is represented by a resistance $M_d^2I_s^2/K_r$ connected in parallel with the capacitance $(M_d^2I_s^2)/M_rp$.

4. Comparison of equations 20 and 21 reveals that the inertia of the rotor due to the initial velocity ω_{r0} is represented by an initial charge

$\dfrac{M_rE_{20}}{M_d^2I_s^2}$ on the capacitor $\left(\dfrac{M_d^2I_s^2}{M_rp}\right)$ of Figure 4.

D-C Machine Flexibly Connected to Load

It is now possible to extend the analogy to include the case of a flexible coupling compliance D_c of the rotor to a load of inertia M_l and viscous damping K_l. The mechanical circuit is shown in Figure 5.

The differential equations for the motion of the system are

$$\omega_r=\omega_c+\text{[illegible]} \quad (22)$$

where

$$\omega_c=D_c\frac{dT_l}{dt}$$

$$T_l=K_l\omega_l+M\frac{d\omega_l}{dt}$$

In these equations, the suffix *c* refers to the coupling and the suffix *l* to the load.

The subsidiary equation is

$$\omega_r=\left(D_cp+\frac{1}{K_l+M_lp}\right)\mathbf{T}_l-D_cT_{l0}+\frac{M_lT_{l0}}{K_l+M_lp} \quad (23)$$

$$Q_l=Q_{l0}+\int_0^t T_ldt$$

Put $T_l=I_lM_dI_s$ and $\omega_r=\dfrac{E_2}{M_dI_s}$; then

$$\frac{\mathbf{E}_2}{M_d^2I_s^2}=\left(D_cp+\frac{1}{K_l+M_lp}\right)\mathbf{I}_l-D_cI_{l0}+\frac{M_lI_{l0}}{K_l+M_lp} \quad (24)$$

The circuit of Figure 4 may now be extended to include the effects of a load connected flexibly to the rotor circuit, as shown in Figure 6(A). Loss in the spring coupling may be represented by a resistance

$\dfrac{M_d^2I_s^2}{K_c}$ in shunt with the inductance $M_d^2I_s^2D_cp$, as shown dotted.

If an external torque is impressed on the rotor, then electromotive force E_2, which is a measure of the angular velocity of the rotor ω_r, is applied such that current I_2 (torque T_r) is delivered to the rotor and current I_l (torque T_l) is delivered to the load. The voltage between points *A* and *B* represents the velocity of the load which can be different from the velocity of the rotor. If the torque transmitted through the coupling is uniform, then I_l will be a direct current, and points *A* and *C* will be at the same potential and consequently have the same angular velocity.

For some particular applications, it may be more convenient to represent the analogy by the dual of Figure 6(A), as shown in Figure 6(B).

Separately Excited D-C Machine with Motion of Stator Permitted

The matrix equation corresponding to equation 20, but which permits motion of stator as well as rotor, is

$$\begin{bmatrix}\mathbf{E}_r\\ \mathbf{T}_r\\ \mathbf{T}_s\end{bmatrix}=\begin{bmatrix}R_r+L_rp & M_dI_s & M_dI_s\\ -M_dI_s & K_r+M_rp+\dfrac{1}{D_{rs}p} & \dfrac{1}{D_{rs}p}\\ -M_dI_s & \dfrac{1}{D_{rs}p} & K_s+M_sp+\dfrac{1}{D_{rs}p}\end{bmatrix}\times\begin{bmatrix}\mathbf{I}_r\\ \omega_r\\ \omega_s\end{bmatrix}+\begin{bmatrix}-L_rI_{r0}\\ -M_r\omega_{r0}+\dfrac{Q_{rs0}}{D_{rs}p}\\ -M_s\omega_{s0}+\dfrac{Q_{rs0}}{D_{rs}p}\end{bmatrix} \quad (25)$$

where

D_{rs} is the compliance of the spring coupling the rotor to the stator, the suffixes r and s referring to the rotor and stator respectively.

Let

$T_r = I_2 M_d I_s$, $T_s = I_3 M_d I_s$, and

$$\omega_r = \frac{E_2}{M_d I_s},\ \omega_s = \frac{E_3}{M_d I_s}$$

Then matrix equation 25 may be rearranged as

$$\begin{bmatrix} \mathbf{E}_r \\ \mathbf{E}_2 \\ \mathbf{E}_3 \end{bmatrix} = \begin{bmatrix} R_r + L_r p + \frac{M_d^2 I_s^2 (K_r + M_r p)}{Z} + \frac{M_d^2 I_s^2 (K_s + M_s p)}{Z} & \frac{M_d^2 I_s^2 (K_s + M_s p)}{Z} & \frac{M_d^2 I_s^2 (K_r + M_r p)}{Z} \\ \frac{M_d^2 I_s^2 (K_s + M_s p)}{Z} & \frac{M_d^2 I_s^2 \left(K_s + M_s p + \frac{1}{D_{rs} p} \right)}{Z} & -\frac{M_d^2 I_s^2 \frac{1}{D_{rs} p}}{Z} \\ \frac{M_d^2 I_s^2 (K_r + M_r p)}{Z} & -\frac{M_d^2 I_s^2 \frac{1}{D_{rs} p}}{Z} & \frac{M_d^2 I_s^2 \left(K_r + M_r p + \frac{1}{D_{rs} p} \right)}{Z} \end{bmatrix} \begin{bmatrix} \mathbf{I}_r \\ \mathbf{I}_2 \\ \mathbf{I}_3 \end{bmatrix} + \begin{bmatrix} M_d I_s \left(\frac{A_{rs} + A_{sr}}{Z} \right) - L_r I_{r0} \\ \frac{M_d I_s A_{rs}}{Z} \\ \frac{M_d I_s A_{sr}}{Z} \end{bmatrix} \tag{26}$$

where

$$Z = (K_r + M_r p)(K_s + M_s p) + (K_r + M_r p)\frac{1}{D_{rs} p} + (K_s + M_s p)\frac{1}{D_{rs} p};\ A_{rs} = -\frac{M_s \omega_{s0}}{D_{rs} p} + \left(K_s + M_s p + \frac{1}{D_{rs} p} \right) M_r \omega_{r0} - \frac{Q_{rs0}}{D_{rs} p}(K_s + M_s p);\ \text{and}$$

$$A_{sr} = -\frac{M_r \omega_{r0}}{D_{rs} p} + \left(K_r + M_r p + \frac{1}{D_{rs} p} \right) M_s \omega_{s0} - \frac{Q_{rs0}}{D_{rs} p}(K_r + M_r p)$$

The dual equation is

$$\begin{bmatrix} \mathbf{I}_r \\ \mathbf{I}_2 \\ \mathbf{I}_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{R_r + L_r p} & -\frac{1}{R_r + L_r p} & -\frac{1}{R_r + L_r p} \\ -\frac{1}{R_r + L_r p} & \frac{1}{R_r + L_r p} + \frac{K_r + M_r p + \frac{1}{D_{rs} p}}{M_d^2 I_s^2} & \frac{1}{R_r + L_r p} + \frac{1}{M_d^2 I_s^2 D_{rs} p} \\ -\frac{1}{R_r + L_r p} & \frac{1}{R_r + L_r p} + \frac{1}{M_d^2 I_s^2 D_{rs} p} & \frac{1}{R_r + L_r p} + \frac{K_s + M_s p + \frac{1}{D_{rs} p}}{M_d^2 I_s^2} \end{bmatrix} \begin{bmatrix} \mathbf{E}_r \\ \mathbf{E}_2 \\ \mathbf{E}_3 \end{bmatrix} + \begin{bmatrix} \frac{L_r I_{r0}}{R_r + L_r p} \\ -\frac{L_r I_r}{R_r + L_r p} - \frac{M_r \omega_{r0}}{M_d I_s} + \frac{Q_{rs0}}{M_d I_s D_{rs} p} \\ -\frac{L_r I_{r0}}{R_r + L_r p} - \frac{M_s \omega_{s0}}{M_d I_s} + \frac{Q_{rs0}}{M_d I_s D_{rs} p} \end{bmatrix} \tag{27}$$

The circuits corresponding to equations 26 and 27 are shown in Figure 7.

Figure 8 shows an extension of the circuit of Figure 7(A) to include the effect of a flexibly connected load to the rotor and spring mounting of stator to foundation. The spring coupling between stator and rotor is shown dotted because normally this will not be used, especially with direct current applied.

Time Scales

Time scaling can be chosen to suit the requirements of a particular problem; for example, for an electric circuit model it is often convenient to contract the time scale to 10^{-3} of the original, thus giving decreased times for dynamic response, and permitting recurrent transients to be displayed as a steady pattern on an oscilloscope. On the other hand, for mechanical models, it may be convenient to expand the time scale, thus giving increased times of response, sufficient to make possible direct visual observation and measurements from instant to instant. Both methods have been used to study the problem of dynamic response to disturbances between synchronous machines in a-c networks.

Laboratory Tests

Determination of Circuit Constants

Measurement of rotor resistance R_r and inductance L_r. If the rotor of the machine is locked and the stator rigidly connected to a massive foundation, then points X and Y in Figure 4 should be at the same potential, since $\omega_r = 0$. The current supplied by the voltage source E_r is limited only by the circuit constants of the armature R_r and L_r, which may be determined. Typical values for the machine under tests are

$R_r = 5.0$ ohms and $L_r = 0.24$ henry

Measurement of $M_d I_s$. For a given field current I_s, the generated voltage across the brushes in the quadrature axis is measured as E_r at constant angular velocity ω_r. The value of $M_d I_s$ is determined as E_r/ω_r.

From tests $E_r/\omega_r = 0.8$ volt per radian per second, and since $I_s = 0.15$ ampere, $M_d = 5.33$.

Measurement of the moment of inertia of the rotor M_r, the viscous damping of the armature K_r, and the compliance of rotor spring coupling D_c.* The mechanical constants used in the analogy may be determined by electrical or mechanical tests.

For the electrical test, the rotor is connected through a spring coupling of compliance D_c to the same rigid foundation as the stator, and the circuit diagram for this arrangement is shown in Figure 9.

By applying a low variable frequency sinusoidal alternating voltage source E_r to the rotor circuit, it is possible to demonstrate series and parallel resonance effects. Both impedance and resonant frequencies should be noted for each type of resonance.

The results obtained under steady-state a-c response are:

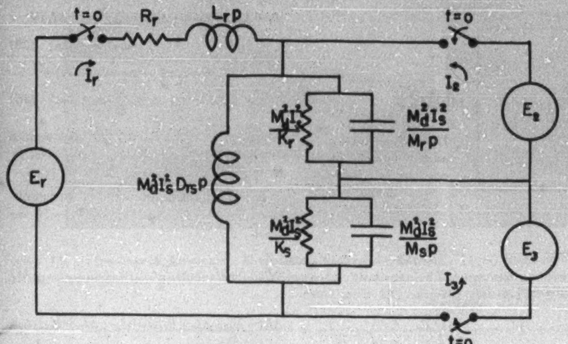
	Impedance in Ohms	Resonant Frequency in Cycles per Second
Parallel resonance	50+j0	0.845
Series resonance	5.0+j0	1.28

Resonant frequencies as calculated from the circuit constants of Figure 12 are 0.840 and 1.27 cycles per second.

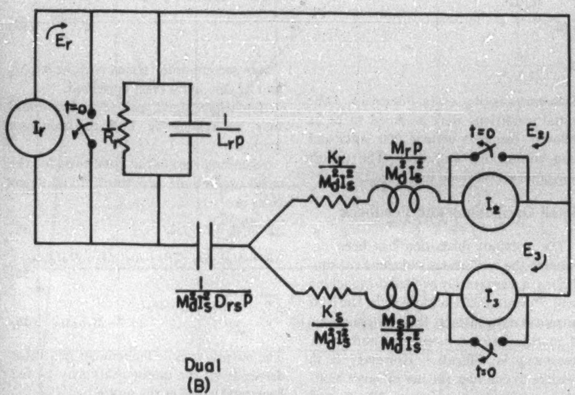
Obviously, the spring coupling D_c may be eliminated in this test, leaving the rotor free when the number of unknown circuit constants is reduced to two, namely K_r and M_r. Under this condition, only series resonance will occur at 0.94 cycle per second.

---

* The viscous damping constants of the armature K_r and the stator K_s are not strictly constant, but in general will be a function of ω_r and ω_s.


t=0
R_r
L_rp
t=0
I_r
I_2
E_2
$\frac{M_d^2I_s^2}{K_r}$
$\frac{M_d^2I_s^2}{M_rp}$
E_r
$M_d^2I_s^2D_{rs}p$
$\frac{M_d^2I_s^2}{K_s}$
$\frac{M_d^2I_s^2}{M_sp}$
E_3
I_3
t=0

Original
(A)


E_r
I_r
t=0
R_r
$\frac{1}{L_rp}$
t=0
E_2
$\frac{K_r}{M_d^2I_s^2}$
$\frac{M_rp}{M_d^2I_s^2}$
I_2
$\frac{1}{M_d^2I_s^2D_{rs}p}$
E_3
I_3
$\frac{K_s}{M_d^2I_s^2}$
$\frac{M_sp}{M_d^2I_s^2}$
t=0

Dual
(B)

Figure 7. Equivalent circuit of a separately excited d-c machine with rotor and stator free to rotate

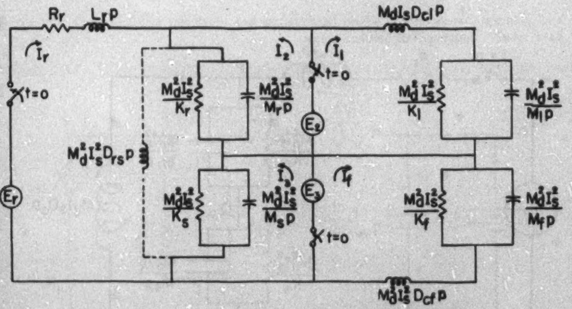

R_r
L_rp
$M_dI_sD_{cl}p$
I_r
t=0
I_2
I_1
t=0
$\frac{M_d^2I_s^2}{K_r}$
$\frac{M_d^2I_s^2}{M_rp}$
E_2
$\frac{M_d^2I_s^2}{K_l}$
$\frac{M_d^2I_s^2}{M_lp}$
$M_d^2I_s^2D_{rs}p$
E_r
I_3
I_f
$\frac{M_d^2I_s^2}{K_s}$
$\frac{M_d^2I_s^2}{M_sp}$
E_3
$\frac{M_d^2I_s^2}{K_f}$
$\frac{M_d^2I_s^2}{M_fp}$
t=0
$M_d^2I_s^2D_{cf}p$

Figure 8. The addition to the circuit of Figure 7 of a flexibly connected load to the rotor and foundation to the stator

s denotes stator, r rotor, c coupling, l load, and f foundation

From these tests, the values of K_r, M_r, and D_c may be calculated. For the machine under test, values of K_r and M_r, as checked by Routin's retardation test, are

$K_r = 0.0142$
$M_r = 0.763$ kilogram-meter^2

$$\frac{M_d^2I_s^2}{K_r} = 45 \text{ ohms}$$

$$\frac{M_r}{M_d^2I_s^2} = 0.119 \text{ farad}$$

It should be noted that the value of $M_r = 0.763$ kilogram meter squared represents the sum of the rotor inertia 0.015 kilogram meter squared and the inertia of a flywheel 0.0613 kilogram meter squared.

Measurement of the moment of inertia M_s and viscous damping K_s of the stator. The spring coupling of the compliance D_c may connect the rotor to the stator which is now free to rotate as shown in Figure 11(A). The appropriate equivalent circuit is shown in Figure 10.

To simplify the circuit, the rotor may be locked to the foundation, the stator being restrained only by the spring. Points P and Q in Figure 10 are then at the same potential since the rotor velocity is zero. The circuit becomes similar to Figure 9 but with K_r and M_r replaced by K_s and M_s.

Again it is possible to demonstrate series and parallel resonant effects and measurements may be taken as previously to determine the values of K_s, M_s, and D_c. If only K_s and M_s are required, then the spring coupling D_c may be eliminated from this test.

The moment of inertia to the stator was measured to be

$M_s = 0.65$ kilogram-meter^2

and

$$\frac{M_s}{M_d^2I_s^2} = 1.02 \text{ farads}$$

The spring compliance $D_c = 0.47$ radian per kilodyne-meter gives the equivalent inductance of

$M_d^2I_s^2D_c = 0.3$ henry

Transient current response to input step function of voltage. For this experiment, the rotor with flywheel rigidly attached was connected through a torsion spring coupling to rigid foundations, the stator being bolted rigidly to the same foundations. The mechanical arrangement is shown in Figure 11(B) except that the stator should be fixed to the foundation. The analogous electric circuit is shown in Figure 9.

With the system at rest, a step function of voltage was applied at time $t = 0$ to the

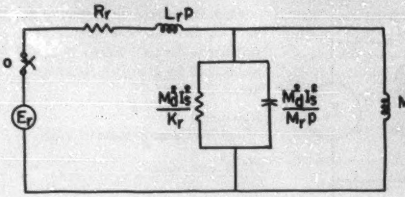


Figure 9. Rotor connected through spring coupling to rigid foundation

armature circuit, the excitation being constant at $I_s=0.15$ ampere. The current response was recorded on an oscillogram, as reproduced in Figure 12 by the dots. Using the measured constants as shown on the circuit diagram in Figure 12, the transient response may be calculated to be

$$i(t)=5.00-5.91\epsilon^{-19\cdot 1t}+1.76\epsilon^{-0\cdot 96t}\cos(5.43t+1.03)$$

The plot of this equation is indicated by the smooth curve in Figure 12. There is a close agreement between calculated curve and the test considering the nonlinear effects which may be present due to coulomb friction, spring, iron losses, and saturation.

DETERMINATION OF FAULT CURRENT

Having determined the constants of the equivalent circuit, it is now possible to calculate the transient fault current of the machine when separately excited as a generator and driven by a constant torque machine through the flexible coupling. In this case, the dual circuit shown in Figure 13 is preferable as constant torque then may be represented by an electromotive force E of a battery.

Initially, the generator is delivering power to a load resistance R_l, when at time $t=0$ the fault is applied. This is equivalent to the opening of the switch in Figure 13 and the transient electromotive force across it is a measure of the fault current delivered by the machine.

The transient voltage appearing on the opening of the switch may be calculated as the voltage which must be impressed in place of the switch to create a steady direct current equal but opposite to that flowing through the switch at time $x=0$.

The transient fault current of the machine is

$$i(t)=E\left[\frac{1}{R_l}+\frac{K_r'}{K_r'R_r+1}-\frac{\left(\alpha-\frac{K_r'}{M_r'}\right)\epsilon^{-\alpha t}}{\alpha L_r(\alpha-\beta)}+\frac{\left(\beta-\frac{K_r'}{M_r'}\right)\epsilon^{-\beta t}}{\beta L_r(\alpha-\beta)}\right]$$

where

$$\alpha,\ \beta=-\frac{1}{2}\left(\frac{R_r}{L_r}+\frac{K_r'}{M_r'}\right)\pm\sqrt{\frac{1}{4}\left(\frac{R_r}{L_r}-\frac{K_r'}{M_r'}\right)^2-\frac{1}{M_r'L_r}},$$

$$K_r'=\frac{K_r}{M_d^2I_s^2},$$

and

$$M_r'=\frac{M_r}{M_d^2I_s^2}$$

This curve is plotted in Figure 14. The initial conditions were assumed to be as follows: terminal voltage 200 volts and load current 6.5 amperes. The circuit constants are as given in Figure 12.

Small Oscillations and Feedback

The previous discussion has been devoted to the oscillations and transient conditions in a separately excited d-c machine with constant field excitation. If the field current is not constant, then the equations of motion become nonlinear and exact treatment is difficult. However, it is possible to consider the use of small oscillations. Replace equations 17, 18, and 19 by

$$E_s+\Delta E_s=R_s(I_s+\Delta I_s)+L_s\frac{d}{dt}(I_s+\Delta I_s) \quad (28)$$

$$E_r+\Delta E_r=R_r(I_r+\Delta I_r)+L_r\frac{d}{dt}(I_r+\Delta I_r)+M_d(I_s+\Delta I_s)(\omega_r+\Delta\omega_r) \quad (29)$$

$$(T_r+\Delta T_r)=-M_d(I_s+\Delta I_s)(I_r+\Delta I_r)+K_r(\omega_r+\Delta\omega_r)+M_r\frac{d}{dt}(\omega_r+\Delta\omega_r) \quad (30)$$

whose ΔE_s, ΔE_r, ΔT_r, ΔI_s, ΔI_r, $\Delta\omega_r$ denote small variations of E_s, E_r, T_r, I_s, I_r, ω_r respectively.

By taking the differences between equations as 28−17, 29−18, and 30−19 we have

$$\Delta E_s=R_s\Delta I_s+L_s\frac{d\Delta I_s}{dt} \quad (31)$$

$$\Delta E_r=M_d\omega_r\Delta I_s+R_r\Delta I_r+L_r\frac{d}{dt}\Delta I_r+M_dI_s\Delta\omega_r \quad (32)$$

$$\Delta T_r=-M_dI_r\Delta I_s-M_dI_s\Delta I_r+K_r\Delta\omega_r+M_r\frac{d\Delta\omega_r}{dt} \quad (33)$$

where second-order terms such as $\Delta I_s\Delta I_r$ and $\Delta I_s\Delta\omega_r$, have been neglected.

Equations 31, 32, and 33 are linear and may be solved by Laplace transform methods.

Assuming zero initial conditions for the small oscillations, the transformed equations are

$$\bar{\Delta}\mathbf{E}_s=(R_s+L_sp)\bar{\Delta}\mathbf{I}_s \quad (34)$$

$$\bar{\Delta}\mathbf{E}_r=M_d\omega_r\bar{\Delta}\boldsymbol{I}_s+(R_r+L_rp)\bar{\Delta}\mathbf{I}_r+M_dI_s\bar{\Delta}\boldsymbol{\omega}_r \quad (35)$$

$$\bar{\Delta}\mathbf{T}_r=-M_dI_r\bar{\Delta}\mathbf{I}_s-M_dI_s\bar{\Delta}\mathbf{I}_r+(K_r+M_rp)\bar{\Delta}\boldsymbol{\omega}_r \quad (36)$$

The output from a tachometer generator driven from the motor shaft may be fed back to the field of the motor.

Suppose that the feedback takes the form

$$\Delta\boldsymbol{E}_s=\pm f(p)\bar{\Delta}\boldsymbol{\omega}_r=\pm(K''p+K')\bar{\Delta}\boldsymbol{\omega}_r \quad (37)$$

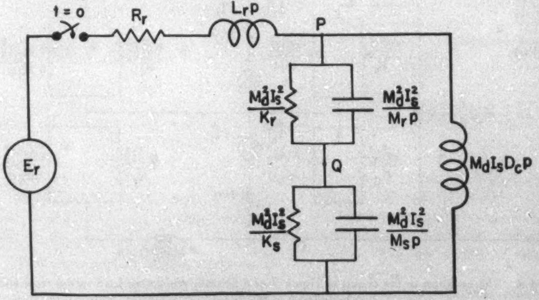


Figure 10. Spring coupling of rotor to stator—stator free

Figure 11. Experimental arrangement of separately excited d-c machine. 11(A) Rotor spring-coupled to stator, stator free. 11(B) Stator free, rotor spring-coupled to foundation, tachometer generator direct-coupled to rotor, and position indicator (Selsyn) friction driven from rotor. Machine rated 220 volts, 6.5 amperes, 2,300 rpm

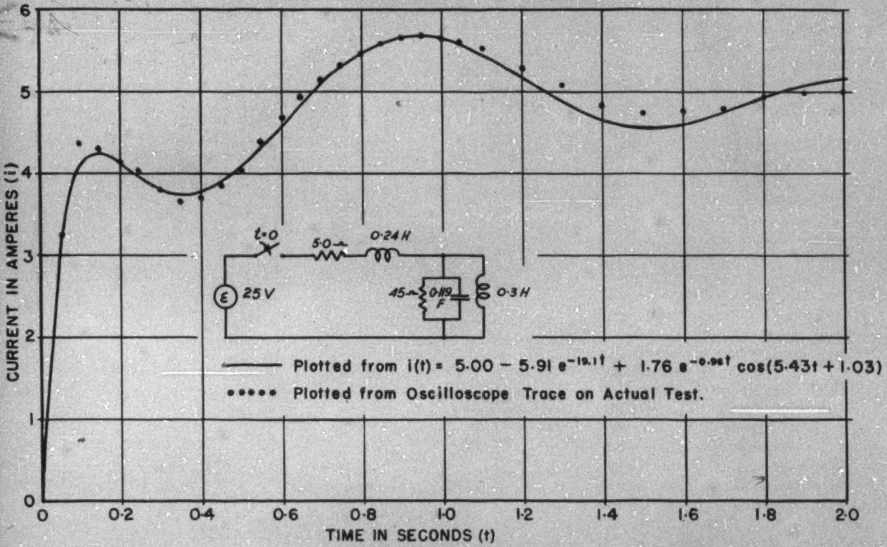


Figure 12. Comparison of calculated and experimental results of rotor current response to a unit step function of magnitude 25 volts. Rotor spring-coupled to foundation. Stator fixed

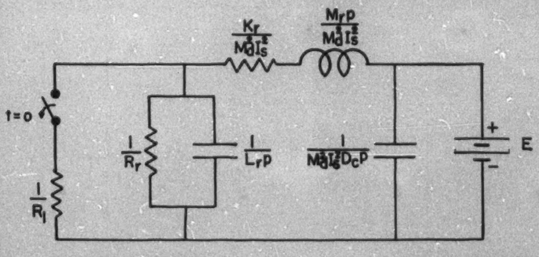

$\frac{K_r}{M_d^2 I_s^2}$
$\frac{M_r p}{M_d^2 I_s^2}$
t=0
$\frac{1}{R_r}$
$\frac{1}{L_r p}$
$\frac{1}{M_d^2 I_s^2 D_c p}$
+
E
−
$\frac{1}{R_l}$

Figure 13. Equivalent circuit of fault on separately excited d-c machine with constant torque applied

On substitution of equation 37 in equation 34 we have

$$\bar{\Delta}I_s = \frac{\pm f(p)\Delta\omega_r}{R_s + L_s p}$$

whence equations 35 and 36 become

$$\begin{bmatrix} \bar{\Delta}E_r \\ \\ \bar{\Delta}T_r \end{bmatrix} = \begin{bmatrix} R_r + L_r p & M_d I_s \pm \dfrac{M_d \omega_r f(p)}{R_s + L_s p} \\ \\ -M_d I_s & K_r + M_r p \mp \dfrac{M_d I_r f(p)}{R_s + L_s p} \end{bmatrix} \begin{bmatrix} \bar{\Delta}I_r \\ \\ \bar{\Delta}\omega_r \end{bmatrix} \tag{38}$$

It will be noted that the mechanical impedance has changed from K_r+M_rp to

$$K_r+M_rp \text{ to } K_r+M_rp\mp\frac{M_dI_rf(p)}{R_s+L_sp}.$$

If the time lag in the field circuit is neglected, then the mechanical impedance becomes

$$K_r+M_rp\mp\frac{M_dI_r}{R_s}(K''p+K')$$

indicating that it should be possible to change the effective damping and inertia of the machine by feedback.

By representing torque by current and angular velocity by voltage, the impedance matrix in equation 38 may be made symmetrical and an equivalent circuit may be drawn as previously.

The block diagram representing equations 34, 35, 36, and 37 is shown in Figure 15.

Conclusions

It has been demonstrated that the method of electromechanical analogies may be applied with advantage to the analysis of rotating machinery. However, the machine selected is a simple one in that there is no magnetic coupling of the armature reaction field with the stator windings and the field current is regarded as constant.

Where electromagnetic coupling between rotor and stator occurs along the pole axis, or the field current is not constant as in metadynes, the equivalent circuit may still be developed, but some elements may be variable. When the problem becomes nonlinear it is still possible to develop an equivalent linear electric circuit by considering small oscillations.

One particular application is that of power-factor control by the addition of capacitance to the rotor circuit of a wound-rotor induction motor by means of the well-known Kapp vibrator. The theory developed readily shows that the oscillating armature of a separately excited d-c machine has as an electrical equivalent, a large capacitance. Furthermore, the analogies developed may be useful in the study of machine vibration problems involving the stator or rotor.

By the use of feedback, it should be possible to modify the mechanical constants of the system to vary the time of response and the general performance of the system.

Appendix. Notes on Duality

It has been shown that the voltage-current equation of the R, L, C series circuit, Figure 16(A), takes the form

$$\mathbf{E}=\left(R+Lp+\frac{1}{Cp}\right)\mathbf{I} \qquad (39)$$

The corresponding equation of the R', L', C' parallel circuit, Figure 16(B), is

$$\mathbf{I}'=\left(\frac{1}{R'}+C'p+\frac{1}{L'p}\right)\mathbf{E}' \qquad (40)$$

If all terms in equation 40 are replaced by their duals, that is, R', L', C', E', I', are replaced by $1/R$, C, L, I, E respectively, then the equation for the dual circuit becomes identical with that for the original circuit.

It is convenient to write the equation of the dual circuit in the form

$$\mathbf{I}=\frac{\mathbf{E}}{R+Lp+\dfrac{1}{Cp}} \qquad (41)$$

since I is the magnitude of the voltage impressed on the circuit and E is the magnitude of the current response.

The impedance of the dual circuit is, in general, the reciprocal of the impedance (that is, operational admittance) of the original circuit.

Extending this to complex circuits, the general voltage-current equation may be written in matrix form

$$[E]=[Z][I] \qquad (42)$$

where $[E]$, $[Z]$, and $[I]$ are matrices.

Written in dual form, equation 42 becomes

$$[I]=[Z^{-1}][E] \qquad (43)$$

where

$[I]$ is the impressed voltage matrix
$[E]$ is the impressed current matrix
$[Z^{-1}]$ is the impedance matrix of the dual network

This provides an alternative method of construction of the inverse of a matrix from the dual circuit.

Example:

The matrix equation

$$\begin{bmatrix}\mathbf{E}_1\\ \\ \mathbf{E}_2\end{bmatrix}=\begin{bmatrix}R+Lp+\dfrac{1}{Cp} & \dfrac{1}{Cp}\\ \dfrac{1}{Cp} & \dfrac{1}{Cp}\end{bmatrix}\begin{bmatrix}\mathbf{I}_1\\ \\ \mathbf{I}_2\end{bmatrix}$$

is represented by the circuit of Figure 16(A).

The dual equation constructed from the dual circuit in Figure 16B is

$$\begin{bmatrix}\mathbf{I}_1\\ \\ \mathbf{I}_2\end{bmatrix}=\begin{bmatrix}\dfrac{1}{R+Lp} & -\dfrac{1}{R+Lp}\\ -\dfrac{1}{R+Lp} & \dfrac{1}{R+Lp}+Cp\end{bmatrix}\begin{bmatrix}\mathbf{E}_1\\ \\ \mathbf{E}_2\end{bmatrix}$$

It will be noted that

$$\begin{bmatrix}R+Lp+\dfrac{1}{Cp} & \dfrac{1}{Cp}\\ \dfrac{1}{Cp} & \dfrac{1}{Cp}\end{bmatrix}^{-1}=\begin{bmatrix}\dfrac{1}{R+Lp} & -\dfrac{1}{R+Lp}\\ -\dfrac{1}{R+Lp} & Cp+\dfrac{1}{R+Lp}\end{bmatrix}$$

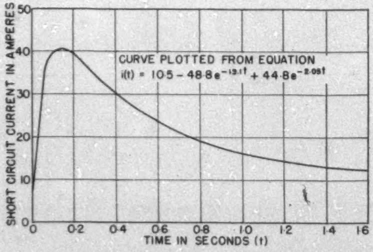


Figure 14 (left). Transient fault current of separately excited d-c machine driven with constant torque

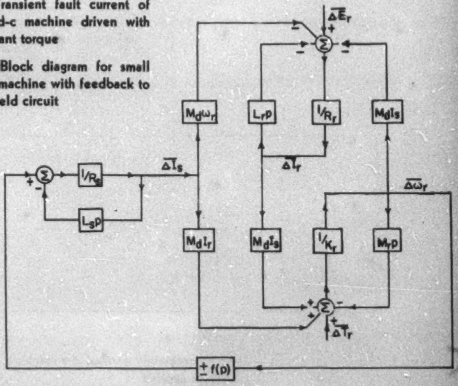


Figure 15 (right). Block diagram for small oscillations of a d-c machine with feedback to the field circuit

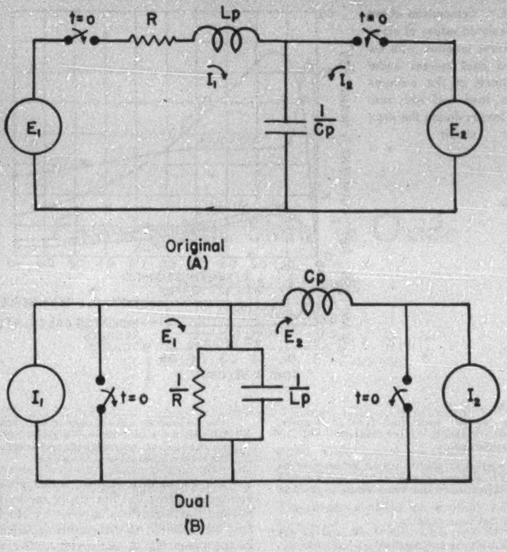


Figure 16. Original and dual resistance-inductance-capacitance series circuit

A rule for the determination of the signs of the elements of the inverse matrix may be found from the fact that the dual of a junction is a mesh, and the dual of Kirchhoff's first law is his second law. There is some similarity between Maxwell's method of stress determination in structural analysis and the method of drawing the dual of a circuit.[2]

Having drawn the dual of a given circuit, it may be necessary to convert some meshes to their equivalent star connections before constructing the impedance matrix of the dual.

References

1. Nouvelles Analogies Electromechaniques, **Gilbert Cahen.** *Bulletin*, Societe Francaise des Electriciens (Paris, France), volume 10, June 1950, pages 288–94.

2. On Methods for the Construction of Networks Dual to Non-Planar Networks, **A. Bloch.** *Proceedings*, Physical Society (London, England), volume 58, 1946, pages 677–94.

3. A New Analogy Between Mechanical and Electrical Systems, **F. A. Firestone.** *Journal*, Acoustical Society of America (New York, N. Y.), volume 4, 1932–3, page 249.

4. Electromechanical Analogies and Their Use for the Analysis of Mechanical and Electromechanical Systems, **A. Bloch.** *Journal*, Institution of Electrical Engineers (London, England), volume 92, part I, 1945, pages 157–69.

5. Transients in Linear Systems (book), **M. F. Gardner, J. L. Barnes.** John Wiley and Sons, Inc., New York, N. Y., volume 1, 1942.

6. Notes on Electromechanical Equivalence, **H. Jefferson.** *Wireless Engineer* (Madison, Wisc.), volume 21, 1944, pages 563–70.

7. Electromechanical Transducers and Wave Filters (book), **W. P. Mason.** D. Van Nostrand Company, Inc., New York, N. Y., 1942.

8. Dynamic Analogies (book), **H. T. Olson.** D. Van Nostrand Company, Inc., New York, N. Y., 1942.

Discussion

John Cybulski (Naval Research Laboratory, Washington, D. C.): I would like to thank the authors for introducing me to an application of the matrix form of equation writing. It appears that some simplification results from this procedure in the final drafting of an equivalent circuit.

Although the authors indicate that the appropriate analogy is the inverse as regards initial representation of the mechanical portion of an electromagnetic transducer, the reasons for this choice are not evident and some discussion is invited for the specific case in the paper.

The following comments and questions are offered in the interest of clarifying certain points in the paper:

1. How good is the assumption that the stator current I_s, remains constant during transient variation of the rotor current I_r? Specifically it would be desirable to check the actual variation of I_s as a function of time experimentally under the test conditions outlined in the section of the paper under "Laboratory Tests" entitled "Transient Current Response to Input Step Function of Voltage," for which the equation for I_r has been written and checked oscillographically. Practically, there is no d-c shunt machine where there are no mutual effects reflected into the stator winding.

2. In the section devoted to "Determination of Circuit Constants" it is noticed that the effects of initial saturation provided by either I_s or I_r were not taken into account in many of the measurements. Does this mean that the constants involved are not dependent upon the magnitude of I_s or I_r? For instance, in "Measurement of Rotor Resistance R_r and Inductance L_r" it appears that no I_s was used during measurement of R_r and L_r.

Another question of functional interest concerns measurement of these two constants R_r and L_r. This writer envisions two basic ways of doing it: a d-c method utilizing the initial rate of rise of I_r and steady state I_r; and an a-c method of some specific frequency.

Under the section on "Measurement of M_dI_s" the authors evaluated M_d at one value of I. It is expected by this writer that the value of M_d would be dependent on I_s and the initial load current I_r.

Further clarification is desired for this part of the paper during determination of constants M_r, K_r, D_c, M_s, and K_s. It was not made clear whether during these measurements the value of I_r or I_s was kept constant at a specific value.

3. In the listing of the existing nonlinear effects during actual tests made under the section on "Transient Current Response to Input Step Function of Voltage," the spring is included. Does this mean the spring under test has been worked past the limitation imposed by Hooke's Law?

4. The calculation of fault current of d-c machines is an important problem today. It would be very interesting to have the results of an actual short circuit of the subject machine under the same initial conditions of section 7.2 for a comparison with the calculated function. Such a trial would also serve to test the assumption of constancy of stator excitation if it too was recorded oscillographically.

R. E. Vowels and **W. G. Forte:** We are pleased to have the discussion of John Cybulski as he has raised some important practical points.

The matrix algebra has been found advantageous in machine analysis, although the full power of the method is not demonstrated in this paper as the equations are regarded as simple from the matrix point of view. Much work remains to be done in connection with the realization of networks from impedance matrices. In the case of the electromechanical transducer utilizing electromagnetic forces, the impedance matrix (see equation 13 of paper)

$$[Z]=\begin{bmatrix} Z_{11} & Z_{12} \\ -Z_{12} & Z_{22} \end{bmatrix} \quad (1)$$

is skew symmetric and the mutual coupling would be unsymmetric. By the use of the inverse analogy in the form of the transformation $T_2=I_2a$ and $\omega_2=E_2/a$ for the mechanical portion of the circuit, the impedance matrix becomes

$$[Z'] = \begin{bmatrix} Z_{11} + \dfrac{Z_{12}^2}{a} & \dfrac{aZ_{12}}{Z_{22}} \\ \dfrac{aZ_{12}}{Z_{22}} & \dfrac{a^2}{Z_{22}} \end{bmatrix} \qquad (2)$$

which is symmetric and can be realized very simply.

It is worth mentioning that for an electromechanical transducer utilizing electrostatic forces, the impedance matrix corresponding to equation 1 is symmetric and the direct analogy would be appropriate.

In reply to the question raised by Mr. Cybulski we comment as follows:

1. The assumption that the stator current I_s remains constant during transient variation of the armature current I_r is quite satisfactory provided that the armature current is kept within the rating of the machine. When saturation of the magnetic circuit is encounted, then there will be mutual effects reflected into the stator winding and the accuracy of the results will depend upon the degree of saturation. No consideration has been given in the theory to nonlinear effects, but there is much scope for the use of electric analogues containing nonlinear elements.

For test conditions outlined under the section on "Transient Current Response to Input Step Function of Voltage," the variation in the stator current I_s was less than 3 per cent.

Under conditions of sinusoidal supply to the rotor, the variation in stator current was still less than 3 per cent provided the rotor current I_r was kept within the rated full load value of ±6.5 amperes peak.

2. A summary of some of the experimental data obtained from the 220-volt 6.5-ampere 2,300 rpm d-c machine with two exciting poles, 1 interpole, and brushes on the neutral axis, follows:

(a) Mechanical constants from mechanical measurements.

(1) Moment of inertia of flywheel, a disc of mild steel attached to the rotor shaft and computed from weight and dimensions as 208 pound-inches^1 or 0.0613 kilogram-meter^2.

(2) Torsional spring of helical form, 18.9 pound-inches per radian or a compliance of 0.47 radian per kilodyne-meter.

(3) Moment of inertia of rotor and flywheel computed from the observed natural frequency (0.845 cycle per second) of torsional oscillation of the rotor and flywheel with spring, 0.0763 kilogram-meter^2.

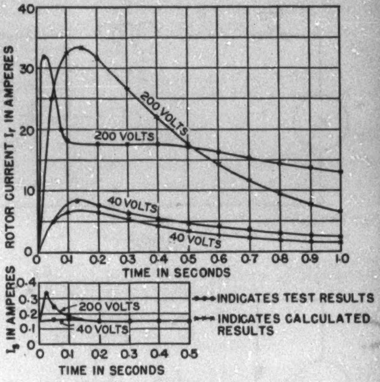
(4) Moment of inertia of rotor only is 0.015 kilogram-meter^2.

(5) Natural frequency of spring and rotor only as measured was 1.90 cycles per second, which checks exactly with the computed value of 1.90 cycles per second from items (2) and (4).

(6) Moment of inertia of stator only was computed from the observed natural frequency (0.895 cycle per second) of torsional oscillation when the rotor and flywheel were spring-coupled to the stator, which was also free to rotate. Using the result of item (3), the moment of inertia of the stator is 0.65 kilogram-meter^2.

(b) The electromechanical constant was obtained from the no-load voltage and speed with measurement at normal excitation $I_s=0.15$ ampere. $M_dI_s=0.8$. Naturally this constant will be affected by saturation.

Figure 1. Comparison of test and calculated values of armature current together with the observed field current under short circuit of the armature from no load and with zero driving torque during the short circuit



(c) The equivalent electric circuit values may be computed from items (a) and (b) here, and it should be pointed out that these values relate to $I_s=0.15$ ampere.

(d) Electrical measurements.

(1) The series inductance of the rotor L_r was measured by several methods:

(1a) Apply alternating current to the rotor with the field current $I_s=0.15$ ampere, and vary the frequency until resonance occurs. For the rotor only (flywheel and spring detached), $2\pi f=13.13$ radians per second, and assuming $M_r'=0.0231$, then L_r 0.25*H*.

(1b) Repeat (1a) but with flywheel attached, then at resonance, $2\pi f=6.0$ radians per second and $L_r=0.23H$.

(1c) Other circuit combinations were tested including the clamping of the rotor to the foundation, again using alternating current at low frequency on the rotor and a constant value of $I_s=0.15$. Generally it is considered preferable, for transient calculations, to measure the inductance by alternating current of a frequency near the main frequencies present in the transient, and in this case the stator current was maintained at normal excitation. A mean value of $L_r=0.24H$ at $I_s=0.15$ ampere was adopted.

(2) Series resistance of rotor R_r: d-c values of R_r varied between 4.5 and 5.0 ohms at rated current and temperature. A-C values of R_r at the resonant frequencies under (1a) and (1b) varied between 5.0 and 5.5 ohms. A mean value of 5.0 ohms was adopted.

(3) Measurements were made under steady-state a-c conditions or transient d-c conditions, avoiding excessive armature currents and spring displacements. A high quality cathode-ray oscilloscope with d-c amplifiers was used throughout and this enabled a constant check to be maintained on possible saturation effects, excessive amounts of unwanted coupling, and so forth. The a-c supply was obtained from a commutator machine having excellent wave form, a low impedance, and frequency stability over the range 0.2 to 30 cycles per second. The field supply was obtained from a separate 400-volt d-c source through a series rheostat which absorbed about 200 volts and maintained the current constant at 0.15 ampere.

3. The listing of the spring under possible nonlinear effects did not imply oscillations beyond the limitations imposed by Hooke's law, but referred to difficulties experienced in the mounting of the spring. The nonlinearity was found finally to be very small. The close agreement between test and calculated transient currents in Figure 12 of the paper indicates that saturation effects were small.

4. The methods presented in this paper were not intended primarily for the calculation of fault current, but it is apparent that Mr. Cybulski is interested in this problem. Obviously saturation will complicate the calculation, and since no allowance was made for this, we did not expect close agreement of the test values and the calculated curve of Figure 14. In the short time available, it was found difficult to check this curve as a constant torque driving source was required. The nearest approach was to use zero driving torque with initial open circuit electromotive forces of 200 volts and 40 volts, and to short-circuit the machine from the open circuit condition. The test and calculated short-circuit curves are shown in Figure 1 of the discussion. The effect of saturation is apparent in the 200-volt test, although peak values correspond. Coupling between rotor and stator is high during the first 0.2 second. The shapes of the calculated and test curves agree in the 40-volt test, and there is little mutual coupling (less than 7 per cent variation in I_s) between rotor and stator.

Characteristic Shifts in Oxide Cathode Tubes

WILLIAM P. BARTLEY
NONMEMBER AIEE

JOHN E. WHITE
MEMBER AIEE

THE stability of control characteristics in electronic tubes is of continuing interest to users and manufacturers alike. This quality is one phase of a broader interest in high quality tube performance evidenced by recent conferences on tubes for instrumentation and computers and by growing demands of the airlines and the armed forces for more reliable tubes.

A factor in this stability which has received a large share of attention in the past few years is the occurrence of a high resistance layer at the interface between base metal and emissive oxide in oxide-cathode tubes.[1–8] There is some evidence[1] that oxide cathode tubes can last almost indefinitely under favorable conditions (20-year lives are quoted), as far as emission is concerned; but practical end-of-life will generally occur long before emission is lost due to shifts in characteristics, in spite of carefully controlled operating conditions.

Interface compounds which cause injurious resistive layers are thought generally to be produced by the chemical reaction of impurities in the cathode base nickel with the emissive coating. Thus Rooksby[2] identified an interface of $BaO.Al_2O_3$ on an aluminum-bearing base metal; Ba_2SiO_4 on a silicon-bearing base metal; and Ba_2TiO_4 on a titanium-bearing base metal. Our tests have indicated interface compounds of Si, Ti, and Mg on "normal" commercial nickel base metals after a period of service.

Circuit-wise, the interface layer acts very much like a resistor shunted by a small capacitance in series with the cathode of the tube. It is thus capable of shifting the static control characteristics. The presence of the capacitance adds distortion of the output to a general shift of characteristics, in computer and other applications where short pulses are used[3,4] and makes the transconductance of the tube frequency dependent.[5] In applications where high peak emission is required, high coating or interface resistance will reduce the available emission and favor destruction of the emissive coating by sparking.[6]

In applications where a tube is to operate at frequencies below about 10^5 cycles per second the capacitance of the interface layer is not important.[5] The low-frequency characteristics do shift, especially during early life, and it is to be expected that part of this effect is due to development of a resistive interface layer. It will be shown in this paper that other factors also are of importance in this shift. A principal objective of the paper is to show what factors contribute to the shift of tube characteristics with time and how they combine to produce the overall effect observed.

Measurement of actual values of interface resistance and capacitance has been carried out using a method developed by W. U. Shipley,* General Electric Receiving Tube Department. Use of this method has made it possible to fit interface impedance into the general framework of factors which cause tube characteristics to change and also to study factors influencing production of the resistive layer. The layer is found to form rapidly on "active" base metals (nickels with large reducing impurity content) but has not been considered a problem on passive base metals (nickel as pure as available). The resistance of the layer is shown to vary with temperature and to exhibit certain transient properties.

Life test data was taken from portions of a broader statistical study being carried out by L. U. Hamvas. His tubes, like those used for our characteristic studies, were *6SN7-GT*. These tubes consist of two complete triodes, independent but for heater connections, in one envelope. Confirmatory tests were made on *6J5GT* tubes.

Interface Impedance Studies

METHOD OF MEASUREMENT

If a square wave signal of several microseconds period is applied to the grid of a tube having interface resistance shunted by capacitance, the output is peaked at the leading edge and tapers off with the *RC* (resistance-capacitance) time constant of the interface impedance. Tubes are tested for the existence of interface layers by superimposing the output signal upon the square-wave input signal on a scope face, using electronic switching. Tubes without interface layers are easily chosen by flat-top output traces in this test. For measurement of interface impedance, the outputs of a reference tube without interface impedance and of a tube under investigation are superimposed. The square-wave input is the same for both tubes. Matching of the outputs is accomplished by adding resistance and capacitance as required in the cathode circuit of the reference tube. The values of parallel *R* and *C* required for matching are considered to be the actual *R* and *C* values of the interface layer in the tube under test.

PROPERTIES

Waymouth[3] found that some tubes with Ba_2SiO_4 between base metal and oxide gave evidence of interface impedance and others did not. This fact was interpreted as indicating an "activation"

Paper 52-53, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted October 19, 1951; made available for printing December 10, 1951.

WILLIAM P. BARTLEY and JOHN E. WHITE are both with the General Electric Company, Schenectady, N. Y.

*W. U. Shipley is with the General Electric Company.

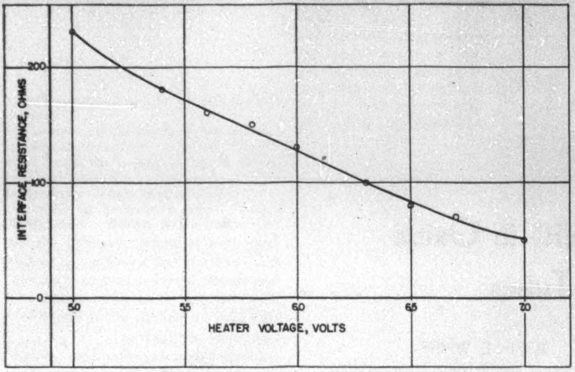


Figure 1. Interface resistance as a function of heater voltage during measurement

of the orthosilicate by excess barium according to current theories of impurity semiconductors. Such behavior emphasizes the need for caution in interpretation of results: a resistance measured as, say, 50 ohms might be different when measured later and might, indeed, change during the very process of measurement. Some of our studies on the properties of these impedances will now be described.

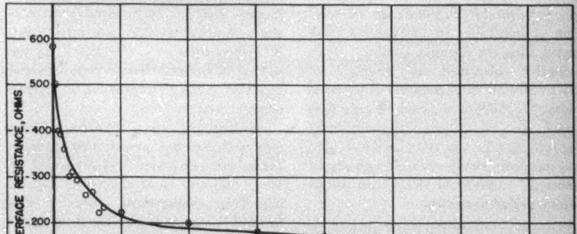
Since the materials considered are probably semiconductors, it is to be expected that their resistances will fall as temperature rises. This effect is shown in Figure 1, plotted against heater voltage. It is clear that heater voltage must be accurately known, for reliable resistance measurements. All such measurements reported here, except for Figure 1, were carried out at the rated heater voltage of 6.3.

Impedance measurements quickly taken were found independent of plate current during measurement, for a reasonable spread on either side of the tube rating of 9 milliamperes. However, continuing plate current is likely to "activate" the layer as mentioned above, giving a progressive reduction in resistance. This effect is shown in Figure 2. Continuing plate current of 7 milliamperes brings an initial resistance of 580 ohms down to a final value of 160 ohms in three hours. Raising the plate current to 10 milliamperes then causes a further drop in resistance to 120 ohms.

If the resistance has been brought to a steady value by a certain plate current which is then cut off, the resistance will rise to a new steady value. The old value will reappear on reapplication of the former plate current. This is shown in Figure 3.

The effect just described and shown in Figure 3 also was observed in tests on some production tubes of a different type. These tubes had been operating continuously at rated current for some time, and were taken off for transfer to another location, where interface tests were made. The tubes showed interface impedance for a very short period only, and then fell as in the dotted curve in the figure. This is a significant result for tubes to be used in standby service, for the tube transconductance measured on a test set in the usual manner would be normal, corresponding to equilibrium or long-time current application in Figure 3. After a period of idleness, however, the transconductance would drop, due to the "hump" effect shown in this figure.

Figure 2. Decay of interface resistance during operation at constant plate current. Upper curve run at 7 milliamperes, followed by lower curve at 10 milliamperes



LIFE TESTS

Life tests have been carried out on 288 tubes (or 576 sections of these double triodes) manufactured in a single batch to reduce uncontrolled variables to a minimum. Three types of cathode base metals were used, classified in the usual manner as "passive", "normal" and "active" according to the quantity of reducing impurities in the nickel. The tubes were divided into groups with different heater voltages in each, some with plate and grid voltages fixed to give rated plate current for an average tube, while others were biased beyond cut-off and carried essentially zero plate current. The curves quoted below give averages for at least six tubes under each condition.

Figure 4 shows the building up of interface resistance with time of three groups of tubes operating cut off. It is seen that the resistance builds up faster in tubes with hotter cathodes. Tests like those of Figure 4, but with grid bias adjusted for rated plate current, are shown in Figure 5. A comparison of these two figures shows clearly that interface resistance builds up faster in tubes operated without plate current.

Figures 4 and 5 both give data on active cathode tubes. Since the upper curve tubes are already beyond useful life, the curves have been terminated at 1,300 hours. No significant resistances had developed in either passive or normal cathode tubes under any conditions, within the period covered by the curves.

These life tests have now reached 3,000 hours. Even by this time, neither normal nor passive cathodes developed signifi-

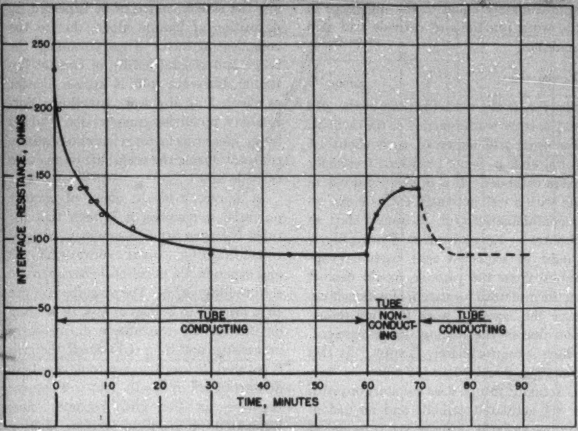


Figure 3. Transient character of interface resistance as induced by plate current changes

cant signs of interface impedance, with or without plate current, at rated heater voltage. Of tubes operated at 7.5 volts on the heater and carrying plate current, those with normal cathodes showed a significant impedance at less than 3,000 hours while those with passive cathodes did not. Both passive and normal cathode tubes developed interface resistance at less than 3,000 hours when operated cut off at 7.5 volts on the heater, the resistance appearing somewhat earlier on the normal cathode tubes.

One may, with caution proper to the limited scope of the tests, draw the following conclusions from the results just quoted:

1. In confirmation of work reported elsewhere, interface impedance forms preferentially in active cathode tubes, and faster when the cathode is run at above normal temperatures and with plate current off.

2. Active cathode tubes do develop enough interface impedance to appreciably affect performance within a few hundred hours, even without abnormal cathode heating and though carrying rated plate current.

3. Either normal or passive cathodes appear satisfactory for 3,000-hour tubes regardless of loading conditions, provided the heater voltage is kept at rated value (cathode temperature at or below 800 degrees centigrade).

4. At cathode temperatures above approximately 800 degrees centigrade, either normal or passive cathodes may develop appreciable interface impedance in less than 3,000 hours, especially if the tubes operate part of the time hot but cut off.

5. The improved behavior of passive cathodes over that of normal cathodes makes it seem possible that still further reduction of certain impurities in the base nickel might entirely eliminate the interface problem.

Shifts in Control Characteristics

Dynamic Characteristics

As mentioned earlier, interface capacitance affects only the dynamic, not the static, tube characteristics. Raudorf[5] gives the following equation for transconductance as a function of the interface resistance, R and capacitance, C:

$$G_m = G_m^0 \sqrt{\frac{1+\omega^2R^2C^2}{(1+RG_m^0)^2+\omega^2R^2C^2}} \qquad (1)$$

where G_m^0 is the transconductance of the tube with no interface impedance and ω is the radians per second frequency of operation.

To indicate the magnitudes involved, we submit Figure 6, showing the variation of measured capacitance as a function of heater voltage, as taken on the same tube used in obtaining Figure 1. For example, if $R=150$ ohms and $C=3\times10^{-3}$ microfarad, a tube with $G_m^0=2.60\times10^{-3}$ mho and operating at 10^6 cycles per

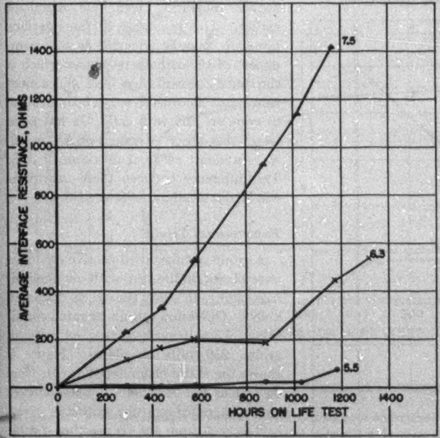
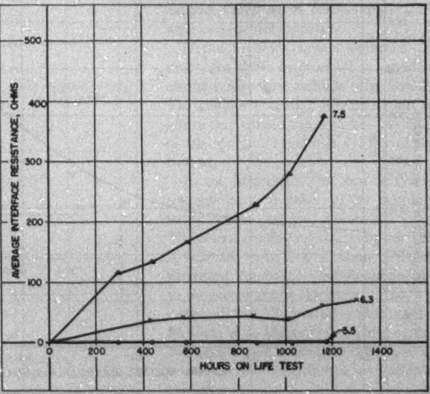


Figure 4. Interface resistance build-up under cut-off conditions. Parameter figures are heater voltages maintained during life

Figure 5. Interface resistance build-up while carrying plate current. Heater voltage during life indicated on curves



second will have its transconductance reduced to 2.47×10^{-3} mho. At $\omega=0$, the transconductance of this tube is only 1.87×10^{-3} mho. At infinite frequency, $G_m=G_m{}^0=2.60\times10^{-3}$ mho.

We should emphasize at this point that the shift in frequency response just indicated is by no means the same as the shift of the same tube in time. That is, $G_m{}^0$ is not necessarily the value of the transconductance the tube had before it developed an interface layer. Thus, Raudorf evaluated R and C by use of equation 1, finding that R rose and C reduced as the life of his tubes increased. One may see from his curves, however, that $G_m{}^0$ also changed with time.

We shall approach the problem from the other direction, in what follows, by study of the static tube characteristics. The slope of these curves may be used to determine G_m for $\omega=0$. From this and the value of R, $G_m{}^0$ may be obtained, using equation 1.

STATIC CHARACTERISTICS

It is demonstrated in the Appendix of this paper that the only factors expected to produce shifts in the static characteristics of tubes like those studied here, are contact potential shifts and cathode coating or interface resistance shifts. We shall consider these separately.

Contact Potential

The contact potential difference of importance, V_c, is that between cathode and grid. Plate potential shifts of the same magnitude are negligible compared to the much larger applied plate voltage. V_c is approximately equal to the difference in the work functions of cathode and grid materials,

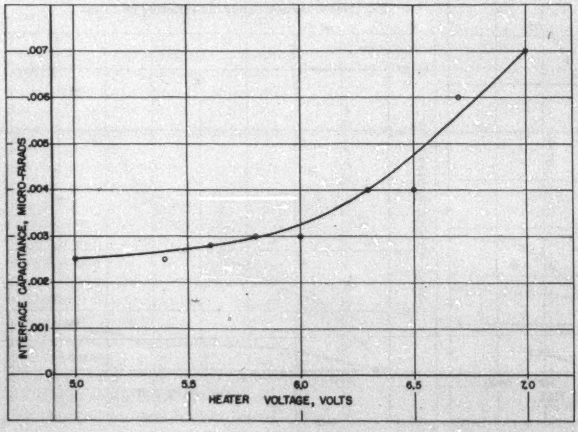
$$V_c=\phi_g-\phi_k \qquad (11)$$

where ϕ_g is the work function of the grid and ϕ_k is the work function of the cathode. For usual grid materials, ϕ_g is about 4.5 volts, while ϕ_k is not far from 1.5 volts for oxide cathodes. If a tube started out in life with a well-activated cathode and an uncontaminated grid, V_c would start at about $4.5-1.5=3.0$ volts. During life, it would be expected that barium evaporated from the cathode would deposit on the grid until formation of a monolayer gave the expected work function reduction due to the barium dipole moment, which is approximately 3 volts. At this time, $V_c=0$ and $\Delta V_c=-3$ volts.

Actually, a tube does not start out with a well-activated cathode and an uncontaminated grid. Since the process of cathode breakdown and activation evaporates barium, the grid is bound to be somewhat contaminated at start of life. The fact that oxygen is released from the cathode during activation and during life also exerts an influence. Further complication is added by the fact that various parts of the grid are at different angles and distances from the cathode and contaminate at different rates. Finally, we have the fact that the work function of a metal like tungsten or nickel goes through a minimum at the time the surface is just covered by a monolayer, ϕ_g then rising by about 0.5 volt as the barium layer thickens.

While the variation of ϕ_g does not follow the simple course to be expected by deposition of barium alone, due to the complicating factor of gas poisoning, the range and general nature of contact potential difference shift is known, inasfar as effected by grid work function. Fortunately, no further complication is added by ϕ_k, since this factor remains essentially constant during the useful life of an oxide cathode tube.

Figure 6. Interface capacitance as a function of heater voltage during measurement



In summary of the effect of contact potential differences, it is seen that for tubes like those studied, one may expect a tube to start life with an arbitrary $\Delta V_c=0$ and experience a horizontal characteristic shift as life goes on. The magnitude of the shift may go to 3 volts or less, depending on the barium evaporation during early processing, and then reduce. If the grid is covered by less than a monolayer of Ba at start of life ($\Delta V_c=0$), ΔV_c will become negative as the grid becomes contaminated up to a monolayer, making ΔV_g from this cause also negative (equation 10).

Our experiments, to be described later, show that ΔV_c follows the pattern described. While the rates of shift of V_c agree in order of magnitude with known evaporation rates from oxide cathodes, direct numerical comparisons are difficult because of the complexities mentioned.

Resistance Drop

The absolute value of the resistance drop caused by plate current flowing through the cathode oxide cannot be found by studying the tube $I-V_g$ curves. The quantity responsible for shifting the curves is, as shown in equation 10, the change in resistance, ΔR. An absolute value of resistance is measured in the interface measuring technique described earlier. This resistance of the interface layer, R_i, may be regarded as that component of the cathode resistance which is shunted by a small capacitor. Since most tubes start life with $R_i=0$, it is interesting to compare ΔR with ΔR_i. It has been found that observed values of ΔR are always at least as great as those of ΔR_i. The difference between these quantities will be discussed in the next section.

Experimental Results

A group of tubes with passive cathodes were placed on life test with no preliminary aging, to study the effects discussed above. Operation was under rated conditions: 6.3 volts on heaters; -8 volts on grids; 250 volts on plates. Figure 7 shows the static characteristic with time on test as the parameter. These curves are averages for ten tube sections.

It is seen from the curves that a rela-

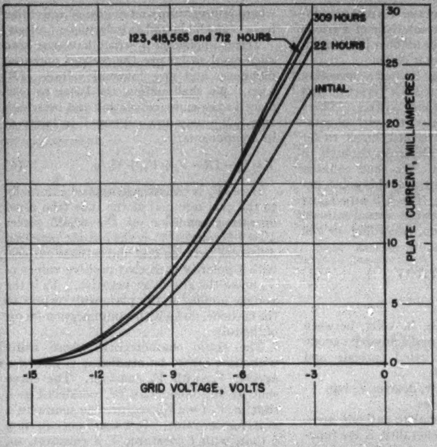


Figure 7. Average static characteristic curve of passive cathode tubes at start of life and after various periods of life in hours indicated by figures on the curves

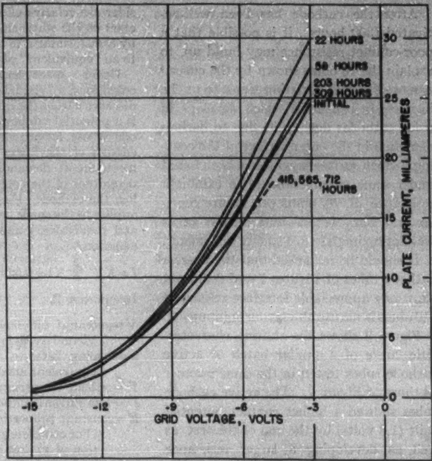


Figure 9. Average static characteristic curve of active cathode tubes, as it varies during life

tively large shift to the left occurs within 22 hours, the shift continuing at a diminishing rate up to 309 hours, after which the direction of shift reverses. This was at first thought to be explained by the building up of a monolayer of Ba on the grid in 309 hours and subsequent reversal of the direction of shift due to reversal of the work function change, as discussed previously. Further analysis has demonstrated the presence of another factor.

A graphical method was developed to distinguish between the effects of resistance and contact potential changes. This method consisted simply of plotting I against V_g on 2/3-power paper for the initial and for a later-in-life characteristic curve, on the same sheet, and extending the resulting straight lines to zero current. The curve separation here is ΔV_c. The value of ΔR is obtained by difference from equation 10 (Appendix) at the highest current measured. It was found that substituting ΔV_c and ΔR determined in this manner into equation 10 and applying the resultant ΔV_g to the initial curve along its entire length, gave a new curve agreeing almost perfectly with the shifted characteristic. Such agreement also was obtained with active cathode tubes, where characteristic shifts were considerably greater.

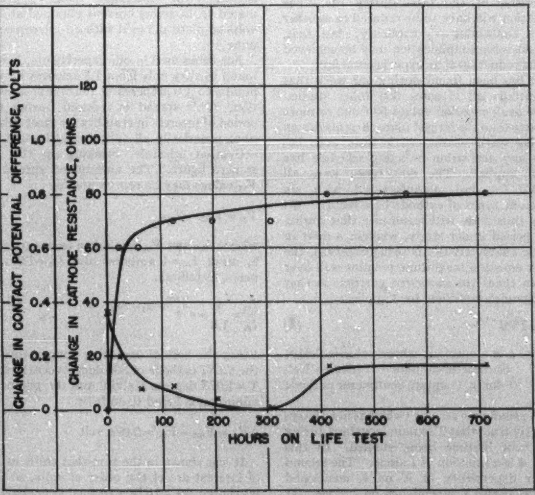


Figure 8. Average changes in contact potential difference and cathode resistance during life of passive cathode tubes. Upper curve, contact potential difference; lower curve, resistance

Figure 8 shows contact potential and resistance shifts determined as described on the ten passive cathode tube sections. The contact potential difference shift requires no further comment, but the shape of the resistance curve is worthy of discussion.

In the first place, the zero of the resistance curve is taken at its minimum value, which occurs at about 300 hours. The actual coating resistance in this tube is expected to be 5 to 10 ohms in a well-activated condition, corresponding here to the zero point on the curve. It is to be expected that the coating resistance was 35 ohms or so higher before plate current had been drawn in the tube because of the relatively low condition of activation at that time.

After the cathode has been well-activated (300 hours), it is possible that a poor-contact resistance may build up, to explain the 16 ohms shown by the curve.[5] A more likely explanation seems to us the increase in resistance which develops in oxide cathodes during life due to barium depletion in the outer surface of the coating.[9] Such surface layers have been found under some circumstances[10] to exhibit a resistance of 20 ohms per square centimeter. The cathode area on our tubes was approximately 3/4 square centimeter.

It should be remarked that the passive cathode tubes of Figures 7 and 8 did not show any appreciable interface resistance (10 ohms is measurable on our equipment).

Figure 9 shows the shifting characteristic curve of a similar batch of active cathode tubes tested in the same manner as those of Figure 7. The active cathode tubes showed a higher contact potential shift (1.4 volts) by the end of the test, as well as developing a larger resistance shift. The average total resistance shift in these tubes went to 112 ohms by the end of the test, at which time the average interface resistance was 86 ohms. The effect of larger resistance is clearly seen in the skewing of the characteristic curve.

Conclusions

In tubes of the type studied, static characteristic shifts with time of service are found to occur primarily due to changing contact potential difference and to changes in resistance associated with the emissive coating. The degree of cathode activation occasions a first-order effect in changing the resistance of the coating but only a negligible effect by influencing the depth of the potential minimum at the virtual cathode. Interface resistance is a major effect after a few hundred hours in active cathode tubes but remains negligible up to around 3,000 hours in tubes with passive cathodes. Interface resistance, where it occurs, is particularly disturbing in square wave or other mixed frequency applications, since the capacitance associated with an interface layer makes transconductance a function of frequency.

Appendix. Static Characteristics: Mathematical Formulation

It has been established[7] that the Child-Langmuir (3/2 power) equation for space-charge limited emission may be applied to a triode if (V_g+V_b/μ) is substituted for plate voltage, using an "equivalent diode" concept. The "equivalent diode" spacing required for the equation is a function of the geometry[7] of the tube. We shall now consider the relative effects of various parameters on the static characteristic of a triode by transforming it in the manner just stated, to an "equivalent" planar diode.

Under space-charge-limited operating conditions, a "virtual cathode" is set up at a distance δ from the cathode surface. There is a potential minimum at the virtual cathode whose magnitude adjusts itself to let enough of the faster electrons through to meet circuit demands. We shall call the magnitude of the potential at this point, below the cathode, V_m. The 3/2 power law applies to the space between virtual cathode and (equivalent) anode, according to the equation.[8]

$$I=KV^{3/2}x^{-2}(1+0.0247T^{1/2}V^{-1/2}) \tag{2}$$

In equation 2,

V = potential difference, in volts, between virtual cathode and equivalent anode

x = spacing between virtual cathode and equivalent anode

T = cathode temperature, degrees Kelvin

I = plate current, amperes

K = constant proportional to cathode area. (For complete generality, K is a function of x; but we shall demonstrate later that for our conditions it is constant).

The parenthetical expression in equation 2 is a correction term worked out by Langmuir to express the effect of electron velocity distribution.

"Internal" spacing, x, is given by

$$x=D-\delta \tag{3}$$

where D is the electrode spacing in the equivalent diode.

With equation 2 as a general guide, it is desired to find what physical parameters produce changes in the static $(I-V_g)$ characteristic of the tube during life. The equation will have to be reduced to another form containing V_g explicitly; but first, considerable simplification may be achieved by introduction of an experimental fact.

It has been found during our work that essentially all of some 600 tubes studied have peak emission values 50 times or more greater than the largest plate currents drawn during static characteristic tests, even before any activation or aging schedule has been applied. The 600 tubes were all *6SN7GT*'s, but manufactured with six different types of cathode base metal. We may thus state with assurance that during the period under study, wherein a drift in static characteristics is being observed, the peak emission magnitude remains well over seven times the measured current, so that Langmuir's equation[8] for δ applies:

$$\delta=aT^{3/4}I^{-1/2} \tag{4}$$

where a is a constant whose value is 2.84×10^{-7} centimeter-ampere$^{1/2}$ (degree-Kelvin)$^{-3/4}$ for a 1 square centimeter cathode area.

We shall now assume (which is not always strictly true) that T remains constant during the tube lifetime being studied. In this case, δ is a function of I alone. The second order dependency of K on δ, mentioned above, is also a dependency on I alone. It now is possible to write equation 2 as

$$I=f(V) \tag{5}$$

where temperature and degree of activation have been eliminated as independent effects.

Terms implied in V which have not been mentioned so far are the contact potential difference and any internal voltage (IR) drop. We shall assume the latter to take place in the emissive coating and interface, only. We may now express V in terms of its components:

$$V=V_g-IR-V_c+V_m+V_b/\mu \tag{6}$$

Here, V_g is the voltage applied externally to the grid terminal of the tube (the usual operating condition of the *6SN7* places $V_g=-8$ volts). V_c is the contact potential difference between grid and cathode surface, with a polarity such that positive values of V_c make the grid more negative. V_b is the voltage applied to the plate with respect to the cathode, and μ is the amplification factor of the tube.

The static characteristic whose shifts with time are to be studied is a plot of I against V_g, with V_b constant. The curve shift at any point may be considered as a shift in V_g $(=\Delta V_g)$ required to maintain I at its former value. It is seen from equation 5 that, with I constant, V is constant, so, from equation 6

$$\Delta V=\Delta V_g-I\Delta R-\Delta V_c+\Delta V_m+V_b\Delta(1/\mu)=0$$

Experimentally, μ is quite constant over the range of general interest, so

$$\Delta V_g=I\Delta R+\Delta V_c-\Delta V_m \tag{7}$$

V_m may be calculated exactly from Boltzmann's equation:[8]

$$I=I_s\epsilon^{-11,600V_m/T} \tag{8}$$

(I_s is the emission obtainable from the cathode with zero electric field at the emitting surface. Our measured values approximated I_s, in giving current obtained at 400 volts on plate and grid with a 1-microsecond pulse.)

For tubes used in our experiments, I_s was found to vary only from 1.8 amperes off the pumps to 4.0 amperes at maximum activation. This spread is reduced during the period of interest in stability for most tubes, since practically all tubes pass through an activation schedule, bringing up the 1.8 ampere figure. The anticipated spread in V_m values may be represented by

$$V_m=V_{m2}-V_{m1} \tag{9}$$

where V_{m2} and V_{m1} are taken from equation 8, using $I_{s2}=4$ amperes and $I_{s1}=1.8$ amperes, as follows:

$$\frac{I_{s2}}{I_{s1}}=\frac{4}{1.8}=\epsilon^{\frac{11,600}{T}(V_{m2}-V_{m1})}$$

Using the normal operating temperature of the *6SN7* cathode of 800 degrees centigrade, T = 1,073 degrees Kelvin, and the potential minimum is found to shift by

$$\Delta V_m=V_{m2}-V_{m1}=0.074 \text{ volt}$$

It was shown in the text that shifts in V_g of interest are of the order of volts, so the maximum ΔV_m to be expected is negligible.

Finally, then, equation 7 becomes

$$\Delta V_g=I\Delta R+\Delta V_c \tag{10}$$

The significance of the considerations above, culminating in equation 10, is that in the tubes studied during this investigation, the only factors of importance in shifting the static characteristics are contact potential difference and internal resistance associated with the emissive coating or its interface, provided the cathode temperature remains constant.

References

1. The Life of Valves with Oxide-Coated Cathodes, **C. C. Eaglesfield.** *Electrical Communication* (New York, N. Y.), volume 28, 1951, pages 95–102.

2. Identification by X-rays of Interface Compounds on Oxide Cathodes, **H. P. Rooksby.** *Nature* (Paris, France), volume 159, 1947, pages 609–10.

3. Deterioration of Oxide-Coated Cathodes under Low Duty-Factor Operation, **J. F. Waymouth, Jr.** *Journal of Applied Physics* (New York, N. Y.), volume 22, 1951, pages 80–86.

4. The Leaky-Condenser Oxide Cathode Interface, **A. Eisenstein.** *Journal of Applied Physics* (New York, N. Y.), volume 22, 1951, pages 138–48.

5. The Change of Mutual Conductance with Frequency Effect of Oxide-Coated Cathodes, **Walter Raudorf.** *Wireless Engineer* (London, England), volume XXVI, 1949, pages 331–37.

6. Some Properties of the Ba_2SiO_4 Oxide Cathode Interface, **A. Eisenstein.** *Journal of Applied Physics* (New York, N. Y.), volume 20, 1949, pages 776–90.

7. Vacuum Tubes, **K. R. Spangenberg.** McGraw-Hill Book Company, New York, N. Y., 1948, page 186.

8. Electrical Discharges in Gases, II, **Irving Langmuir, K. T. Compton.** *Review of Modern Physics* (New York, N. Y.), volume 3, 1931, pages 241–44.

9. A Study of Oxide Cathodes by X-ray Diffraction Methods, **A. Eisenstein.** *Journal of Applied Physics* (New York, N. Y.), volume 17, 1946, pages 434–43.

10. The Potential Distribution in Pulsed Oxide-Coated Cathodes and Its Consequences for the Velocity Distribution of the Emitted Electrons, **R. Loosjes, H. J. Vink, C. G. J. Jansen.** *Journal of Applied Physics* (New York, N. Y.), volume 21, 1950, pages 350–51.

No Discussion

A High-Speed Direct-Writing Oscillograph

I. S. BLUMENTHAL
ASSOCIATE MEMBER AIEE

A. R. ECKELS
ASSOCIATE MEMBER AIEE

THE Oscilloscribe, developed in the Servomechanisms Laboratory of the Department of Electrical Engineering, Yale University, is a very high-speed direct-writing oscillograph which produces an immediate permanent record of voltages having frequency components ranging from direct current to 200 cycles per second. Development of this device was started in February 1947 when comparable devices were not available commercially. Since that time several instruments of this general type have appeared on the market and have found a wide variety of applications.[1]

The Oscilloscribe, however, is markedly superior to previously reported devices in several important respects. This superiority is made possible by the application of position feedback to the problem of the high speed recording voltmeter. Position feedback had been applied previously in recording voltmeters but the fastest of these could not respond beyond 15 cycles per second. The Oscilloscribe is actually a high performance servomechanism whose block diagram is shown in Figure 1.

The basic element of the instrument shown in Figure 2 is a D'Arsonval type of meter movement having no restoring springs. A 3-inch pen is attached to one end of the movement and a single plate capacitor rotor is attached to the other end. The position of the capacitor rotor is translated into a d-c voltage by means of a radio frequency capacity bridge and a detector. The detector output is compared with the applied signal and their difference is amplified and applied to the moving coil. The coil moves until this difference becomes zero. The pen is then held in this position on a paper chart and is caused to draw a continuous line by running the chart at a constant speed. Figure 3 shows the complete instrument in its final form.

A fairly extensive set of requirements were placed on the system at the start of the project. The meeting of these requirements led to many interesting design problems. In order that the reader may fully appreciate the nature of these problems, the system requirements are listed here.

1. Record instantaneously and permanently voltages of a steady or varying nature.
2. Instrument must be portable.
3. Power requirements—110 volts, 60 cycles only.
4. Frequency response—flat within 3 decibels from direct current to 200 cycles per second.
5. Maximum sensitivity—1 millivolt per millimeter.
6. Maximum drift—2 per cent of full scale after warm up.
7. Full scale deflection—±2.5 centimeters at a 3-inch radius.
8. Linearity—within 1 per cent over whole range.
9. Calibration—adjustable in steps from 25 millivolts to 250 volts full scale.
10. Zero position—adjustable to any point on scale.
11. Input impedance—1 megohm on all ranges.
12. Chart drive—three constant speeds.

Any servomechanism is comprised of component functional units, each of which must be designed by keeping in mind the ultimate performance of the entire system. In a high performance servo of the type being described here, certain units become very crucial and must be designed in accordance with very close tolerances while others, requiring less stringent tolerances, can be designed with an eye to minimizing size and cost. These units will now be described in succession, along with certain considerations which affected their design.

The Pen and Pen Motor

The magnetic circuit of the pen motor consists of a permanent magnet with cylindrical pole faces which are separated from a cylindrical iron core by a one-eighth-inch air gap. The moving element consists of a 500-turn coil of number 40 copper wire wound on a nonconducting rectangular frame which is rigidly attached to a 1/16-inch steel shaft. The shaft passes through the hollow center of the stationary iron core and is supported by ball bearings mounted at the top and bottom of the core.

Current is supplied to the coil through two flexible beryllium copper wires. When current passes through the vertical portions of the coil which are situated in the radially directed magnetic field of the cylindrical air gap, a torque is developed which is proportional to the current. The coil is then free to rotate until it encounters sponge rubber stops at ±20 degrees. A high resistance is connected in series with the coil during the brief warm-up period to prevent damaging currents which can result from temporary amplifier unbalance.

The actual writing is done by a 3-inch aluminum pen attached to the top of the shaft. This pen is a modified version of a very light weight commercially available model which ordinarily depends on capillary attraction to draw ink from the well. Since this pen was originally designed to

---

Paper 52-54, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted July 10, 1951; made available for printing December 14, 1951.

I. S. BLUMENTHAL is with the Northrop Aircraft, Inc., Hawthorne, Calif., and A. R. ECKELS is at North Carolina State College, Raleigh, N. C.

The authors wish to acknowledge the continued encouragement, during the course of this development, of Dr. J. L. Bower of North American Aviation, Inc., formerly Associate Professor of Electrical Engineering at Yale University, New Haven, Conn.

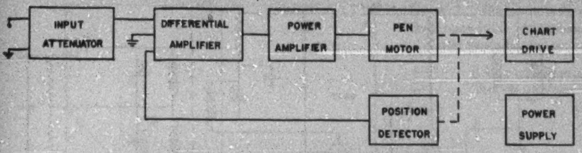


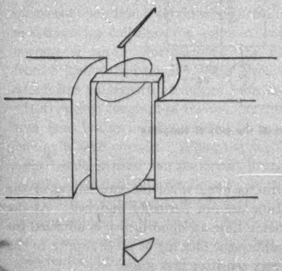
Figure 1. Functional block diagram of the Oscilloscribe

work at writing speeds up to only 100 cycles per second, two modifications were needed. It was found necessary to stiffen the pen considerably in the lateral direction since it resonated like a vibrating reed at about 120 cycles per second. A triangular shaped aluminum truss structure accomplished the desired result at the expense of an increase in moment of inertia. At the same time, this truss provided the type of aerodynamic surface necessary to keep the pen down on the chart whenever the pen is oscillating at very high speeds. In addition, the opening at the pen's tip was enlarged and the inkwell was mounted above the plane of the pen so as to permit a flow of ink sufficient to draw a clear line on the chart even when the pen is oscillating at very high speeds.

Situated at the bottom end of the shaft is a thin aluminum capacitor rotor having a half-inch radius and a 70-degree sectoral angle. This rotor moves in the horizontal plane between two split stator plates. The detector circuit is so designed that the rotor operates at ground potential and thus requires no special insulation from the shaft. To insure that a reliable radio-frequency ground connection always exists, a flexible beryllium copper lead wire is used as a direct ground for the rotor.

The pen and pen motor are obviously very important elements of the system. However, because of the use of position feedback, their precise behavior in this system is much less critical than in comparable spring-restrained devices. In order to see just how their characteristics affect the system behavior, an analysis of the moving system is presented in Appendix I.

Figure 2. Sketch of the basic element of the Oscilloscribe showing the pen, the pen motor, and the capacitor rotor of the position detector



The Position Detector

The position detector is a very critical part of the entire system so far as accuracy is concerned since it must convert the position of the pen into a voltage suitable for comparison with the input voltage. Furthermore, this conversion must be done linearly, with reversible polarity, without appreciable time delay, and without introducing spurious signals. These spurious signals might arise from electrical sources such as drift and hum or they might arise from mechanical imperfections which cause motions of the pen motor shaft other than pure rotation.

A variable capacitor operating with a radio-frequency signal applied was chosen as the transducer because of its small size. Furthermore, when used in a split stator bridge circuit it can result in a detector which is insensitive to vertical or radial motions of the shaft.

The variable capacitor and the demodulating circuit are shown in Figure 4. A 4.3-megacycle signal is generated by a crystal oscillator amplified by a resistance-capacitance coupled triode and a pentode limiter. The pentode supplies an essentially constant current to the tuned primary of the radio-frequency transformer which has an electrostatically shielded secondary. The shielded construction insures that voltages appearing in the secondary arise from induction only. Furthermore, the shield insures a reasonably even distribution of capacity from the secondary to ground so that the apparent ground appears at the midpoint of the secondary.

By grounding the rotor of the transducer and connecting the stators to the ends of the secondary, the apparent ground can be made to shift with the position of the rotor. A high impedance diode circuit connected between one end of the transformer and ground will develop a d-c voltage which depends upon the position of the apparent ground on the transformer secondary. Two such diode circuits connected as shown permit the output to reverse polarity when the rotor swings from positive to negative angles.

A 2-megohm resistor serves to provide a return path for the direct current of both rectifiers without affecting the position of the apparent ground. A careful analysis based upon the assumptions that the distributed capacitance far exceed the maximum capacitance of the rotating capacitor and that the frequency is sufficiently high so that the reactance of the total capacitance to ground from each end the secondary is very much smaller than the effective diode circuit-impedance shows that the output voltage is directly proportional to the angular position of the rotor. Furthermore, under these conditions it can be shown that the detector is essentially insensitive to vertical or radial motions of the shaft.

The Differential Amplifier

The differential amplifier and the power amplifier shown in Figure 5 have numerous advantageous features. When properly designed the balanced type of amplifier used in the first three stages produces an output voltage between the plates of the tubes which is strictly proportional to the difference in the voltages applied to the grids of the tubes. Thus, the difference in the plate voltages of the first stage is proportional to the difference between the input voltage and the position feed-back voltage.

Since the plate voltages of the first

Figure 3 (right). The experimental model of the Oscilloscribe in its final form

stage are both quiescently at a very high d-c potential, their difference cannot be amplified directly by a single triode. This requires that the second stage also be a subtractor, with the plate voltages of the first stage applied to each of its grids. The third stage also operates in this fashion and supplies its output voltage directly to the grids of the balanced power amplifier.

It is evident that, if the tubes are perfectly matched in this type of design, noise in the power supplies, hum from the filaments, or changes of supply voltages will introduce none of the spurious signals which can occur in single ended designs.[2,3] Even when the two halves of each subtractor are not perfectly matched discrimination against common mode signals is very good. Tests made on a *6SC7* twin triode with common cathodes showed that a 10-volt change in B^+ voltage caused a 0.4-volt output signal. Referred to the input, this is equivalent to a spurious signal of 0.01 volt. Since the sensitive amplifiers are connected to power supplies and filament supplies which are themselves well regulated and filtered, it can be seen that the combination is a very simple yet effective method of guarding against drift and noise trouble which can arise in high gain d-c amplifiers.

Provision for quiescently balancing the amplifier or setting the pen to any desired zero position is made by including a potentiometer in one of the coupling networks between the first and second stages. In this position the zeroing arrangement places no special requirements on the regulation of the power supplies.

Attenuation is deliberately introduced between the second and third stages so that phase lead equalizers may be incorporated into the system by simply adding a capacitor on each side of the amplifier. The shunt resistors at this point are potentiometers which are ganged together. This provides a means of adjusting the gain of the amplifier without upsetting the bias arrangements since the cathode of the third stage will follow common changes in its quiescent grid voltages.

The 1-megohm input attenuator is designed in steps of 10 to 1 and 4 to 1. The attenuator switch has a sliding contactor which prevents the grid circuit from opening during the switching process. Resistors of 200,000-ohm capacity are inserted in each grid lead so that the grid circuit resistance will be approximately equal on all ranges. This eliminates shifting of the zero between ranges by equalizing the voltages caused by grid current flow in each triode.

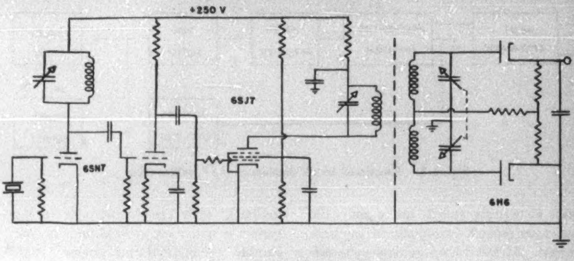


Figure 4. Schematic diagram of the position detector showing location of the split-plate capacitor (dashed line)

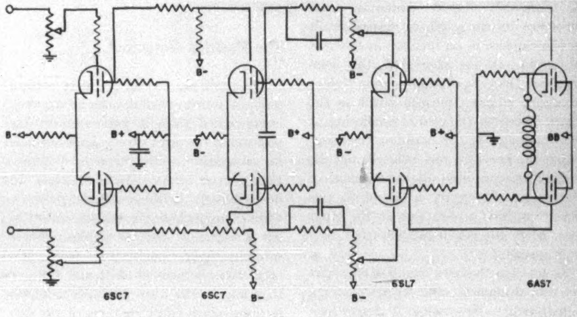


Figure 5. Schematic diagram of the differential amplifier and the power amplifier

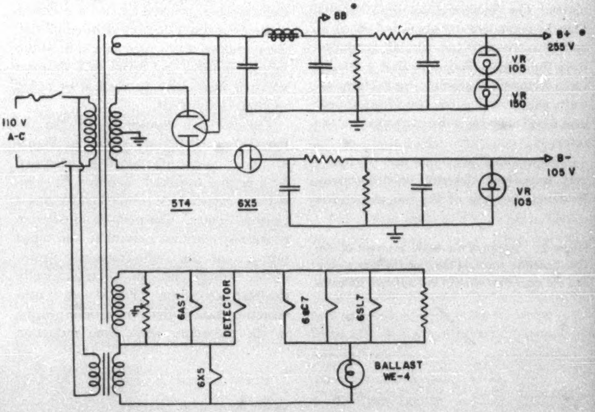


Figure 6. Schematic diagram of the power supplies

The first stage which amplifies signals in the 10-millivolt range is operated at a reduced plate potential to permit the use of a larger cathode resistor, thus making the circuit a more nearly perfect subtractor even when the two halves of the tube are not perfectly matched. At the same time an opportunity is afforded for additional filtering of plate supply to the very sensitive first stage.

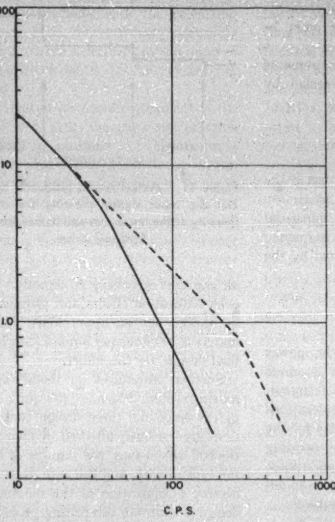


Figure 7. The magnitude of the open-loop transmission ratio versus frequency

——— Before equalization
– – – After equalization

The Power Amplifier

The power amplifier's function is to drive current through the moving coil proportional to the amplified error voltage. This current must be sufficient to develop the peak values of torque required to oscillate the moving system sinusoidally with a prescribed amplitude particularly at the high end of the instrument's frequency range.

The power amplifier which is employed for this purpose is peculiarly suited to the requirements of this instrument. As shown in Figure 5 it is a balanced cathode follower which is capable of supplying a large reversible current to a 2-terminal coil without requiring a B^- supply. This circuit has the particularly advantageous features of requiring no high current negative voltage supply, no center tap on the moving coil, no special filtering or regulation on the B^+ supply, and under quiescent operation it produces no heating in the moving coil since the quiescent coil current is zero. Furthermore, the effective source resistance of this circuit, which has an important influence on the effective time constant of the moving system, is as low as usually can be obtained in single tube circuits. When the *6AS7* twin triode is used in this circuit it can supply ±90 milliamperes linearly to the coil when a potential difference of ±60 volts is applied to the grids. This circuit has a source resistance of 200 ohms and a gain factor of 0.6.

The Power Supply

As will be discussed in later paragraphs, the ability of the system to write with appreciable amplitude is governed by the maximum current available for accelerating the moving system. Since the entire unit was to be portable, a 225-milliampere supply was chosen as the largest practical size. Its circuit is shown in Figure 6.

The power amplifier required 200 milliamperes at 250 volts which, because of the balanced design and the over-all position feedback, did not have to be perfectly filtered or regulated. The output of the capacitor input filter is 385 volts unregulated with a 1.3-volt rms ripple.

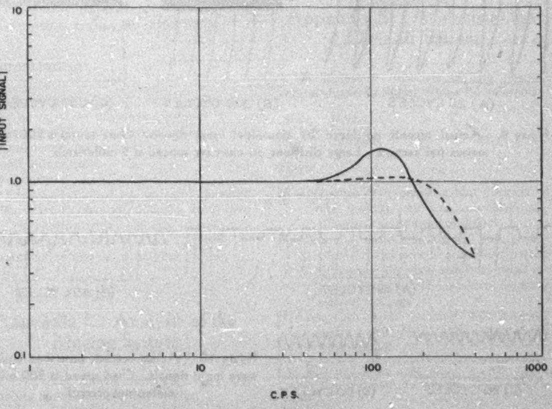
The d-c amplifier required +250-volt and −150-volt regulated supplies while the detector also required +250 volts. The total current required was approximately 20 milliamperes. The critical first stage of the amplifier placed the most stringent requirements on the power supply design since any noise or drift originating in the first stage would be indistinguishable from the applied signals. However, because of the balanced design of the amplifiers, gas tube voltage-regulation together with an additional stage of resistance-capacitance filtering proved adequate for the system. The ripple voltages finally measured were 0.017-volt rms on the +250-volt supply and 0.022-volt rms on the −150-volt supply.

In addition to the foregoing precautions against the introduction of drift signals, a further step was taken against the situation which might arise when changes of filament voltage could introduce different contact potentials in each half of the very critical first stage if its two halves were not perfectly matched. The use of a constant current ballast tube in the filament supply thus eliminates the need for selecting carefully matched twin triodes for the first stage.

A balancing potentiometer connected across the filaments of the first stage reduced the 60-cycle hum level to the point where it causes no observable effect on the written track.

Figure 8. Frequency response for the complete Oscilloscribe

– – – Calculated ——— Tested



The Chart Drive

The chart drive consists of a 75-rpm single-phase motor which drives the chart through a gear box. The gear ratios may be changed by manual adjustment to permit chart speeds of 35, 100, and 200 millimeters per second. The motor pulls the chart along the chart table while the feed roll is friction restrained so that the chart remains taut. The weight of the paper falling freely provides a satisfactory take-up system for the paper coming out of the chart drive.

The 200-millimeter-per-second maximum speed permits each cycle of a 200-cycle-per-second signal to be traced out in a 1-millimeter interval.

Dynamic Performance Requirements and Over-All System Design

The major objectives in the over-all system design were twofold:

1. The frequency response should be essentially flat from direct current to as high a frequency as possible.
2. At the high end of this frequency range the system must be capable of deflecting with an appreciable amplitude.

It is apparent that although item 1 may be the prime objective of the entire design, its realization will be of little consequence if objective 2 is not realized simultaneously.

Actually, it happens that objective 2 imposes the more fundamental requirements on the various functional elements of the system. Therefore, it must be given first consideration here.

The exact nature of the requirements imposed by objective 2 is revealed by equation 3 of Appendix I.

$$i=\frac{J}{K_t}S^2\theta$$

From this equation it follows that if the pen is to deflect linearly with a maximum amplitude θ_m in response to a sinusoidal input of maximum angular frequency ω_m, then the peak current required by the moving coil is given by

$$I_m=\frac{J}{K_t}\omega_m{}^2\theta_m$$

This result suggests that the power supply and the power amplifier be made as large as possible to supply this current. However, size and weight considerations fix the maximum current capacity for the power supply and the corresponding power amplifier. It then becomes a matter of designing the moving system in accord with the following considerations:

1. All of the available current capacity should be usable in a linear manner without encountering voltage limitations.
2. The coil must be capable of carrying the maximum available current without damage to itself.
3. The inertia/torque-constant ratio (J/K_t) should be minimized.

One procedure clearly indicated is to increase the magnetic flux density in the air gap thus increasing K_t directly. However, saturation effects and physical size limitations set an upper limit, approximately 6,500 lines per square inch in the final design, for this effort.

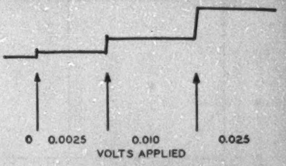


Figure 11. Actual record produced by several d-c input signals showing the over-all linearity of the instrument and its freedom from frictional errors

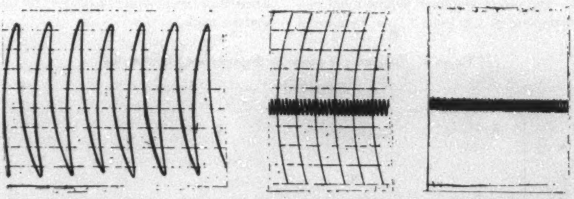

(A) 26 CYCLES (B) 200 CYCLES (C) 250 CYCLES

Figure 9. Actual records produced by sinusoidal input signals. Chart speed is 200 millimeters per second. Large divisions on chart are spaced at 5 millimeters

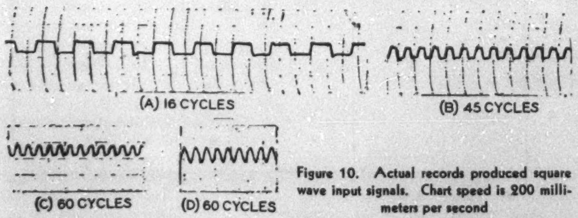

(A) 16 CYCLES (B) 45 CYCLES (C) 60 CYCLES (D) 60 CYCLES

Figure 10. Actual records produced square wave input signals. Chart speed is 200 millimeters per second

Further procedures for modifying the moving system design are not quite so apparent since the three design considerations are variously affected by changes in the coil dimensions, the number of turns, the type of wire, and the wire size. Engineering optimization of the moving coil design is quite an interesting problem in itself. However, a discussion of this problem must be omitted here in the interests of brevity and clarity.

With objective 2 accounted for, the realization of objective 1 may be achieved by the application of standard servomechanism techniques.[6] The quantitative details of this procedure are given in Appendix II. Qualitatively the procedure involves making the open loop transmission ratio appreciably greater than unity from direct current to 200 cycles per second. At the same time however, care must be taken lest the phase shifts around the loop become excessive, with self-oscillation occurring as a result.

The magnitude of the open loop transmission characteristic versus frequency (after initial adjustment of the amplifier gain) is shown in Figure 7. It can be seen that this characteristic is unsatisfactory since it exceeds unity up to only 80 cycles per second. A further increase of amplifier gain would result in instability, however, due to excessive phase shifts. This situation is considerably improved by the insertion of a phase lead equalizer network, resulting in the open loop transmission characteristic shown as a dashed line in Figure 7.

The over-all closed-loop frequency response of the system corresponding to this final open loop characteristic is derived in Appendix II and plotted in Figure 8 as a dashed line. The solid line in Figure 8 is the frequency response obtained by actual test of the final system. Although the actual system response does not agree precisely with the calculated response, it is, nevertheless, essentially uniform, within 4 decibels, over a range of frequencies from 0 to 220 cycles per second.

The reason for the disagreement of the actual test with the response function predicted by equation 11 in Appendix II is simply that this equation does not completely describe the final system. Due to an unfortunate resonance effect in the slender shaft of the moving system, a very pronounced peak occurred in the response between the power amplifier and the detector output at a frequency between 400 and 500 cycles per second. This resonance effect coupled with the increased gain at these frequencies introduced by the phase lead network were sufficient to cause the over-all system to oscillate unstably at a frequency somewhere between 400 and 500 cycles per second. In order to eliminate this oscillation it became necessary to introduce a gain reducing network which had to start becoming effective at about 150 cycles per second so that it could cause the required reduction of gain above 400 cycles per second. Since gain reducing networks inherently introduce phase lag, it is evident that they considerably upset the phase characteristic and sufficiently reduce the system damping so as to cause a resonant peak of approximately 1.5 to occur.

The reduction of the resonant peak is thus a simple matter of a mechanical redesign of the moving system. The final result described here remains, nevertheless, a convincing demonstration of the effectiveness of position feedback in the development of high speed direct-writing oscillographs.

System Performance

Several other aspects of the performance capabilities of this instrument are presented in Figures 9, 10, and 11. Figure 9 shows sinusoidal records at 26, 200, and 250 cycles per second. Calculations based on actual tests indicate that the instrument is capable of reproducing sine waves linearly with a full scale amplitude of 2.5 centimeters for frequencies up to 50 cycles per second. Similar calculations indicate that sinusoids up to at least 150 cycles per second can be reproduced linearly with a peak to peak deflection of 5 millimeters. Evidence that the instrument is indeed capable of clearly observable deflections at these high frequencies is presented in Figures 9(B) and 9(C).

Figure 10 shows the records produced by the instrument in response to square waves. Transient response data can be obtained from the 16-cycle square wave record of Figure 10(A). It can be seen that the pen reaches 1/eth of its final position in less than 2 milliseconds, overshoots slightly, and then settles down completely within 5 milliseconds. Figure 10(B) shows that the record is very definitely a square wave even at 45 cycles per second while the record of Figure 10(C) showing a 60-cycle square wave is neither a sinusoid nor a square wave. The 60-cycle sine wave record of Figure 10(D) is shown by way of contrast to point out the presence of a considerable amount of third harmonic in the 60-cycle square wave record.

The moving element is capable of developing a maximum torque of 2.8×10^5 dyne-centimeters at 0.1 ampere. When the loop is closed an externally applied torque of 1.1×10^4 dyne-centimeters is required to deflect the pen 1 millimeter from its correct position. This high degree of stiffness considerably reduces the effect of viscous or coulomb friction on the pen's motions. Therefore, this system does not suffer the frictional errors which are present in open loop recorders when the signal amplitudes are small. This is clearly demonstrated in Figure 11 where the input is constant but the attenuator setting is changed. This test also demonstrates the over-all linearity between the pen deflection and the voltage applied to the instrument. Drift after warm up has been observed to be less than 1 millimeter.

Because of its high static accuracies and its superior dynamic performance capabilities, it is felt that this type of recorder will find wide application in electrocardiography, electroencephalography and the other rapidly developing fields of electrophysiology, servomechanism research and design, 60-cycle power system study, vibration measurement, nuclear radiation measurement, and many other fields.[7]

Nomenclature

R_a = source resistance of power amplifier
G_a = gain constant of power amplifier
e_a = signal applied to power amplifier
R_c = coil resistance
L = coil inductance
i = coil current
K_t = torque developed per unit coil current
K_v = counter electromotive force developed per-unit angular velocity of coil
θ = coil position
J = moment of inertia of moving system
$R = R_a + R_c$

Appendix I. Analysis of the Moving System

If frictional effects are neglected the following equations may be written:

$$G_a e_a = iR + L_c \frac{di}{dt} + K_v \frac{d\theta}{dt} \tag{1}$$

$$J\frac{d^2\theta}{dt^2} = K_t i \tag{2}$$

In terms of the Laplace trausform these equations yield the following results:

$$i = \frac{J}{K_t} S^2\theta \tag{3}$$

$$\frac{\theta}{e_a} = \frac{G_a/K_v}{S[T_m T_L S^2 + T_m S + 1]} = H_1 \tag{4}$$

where

$$T_m = \frac{RJ}{K_v K_t}$$

$$T_L = \frac{L}{R}$$

The roots of the denominator of equation 4 are

$$S_1 = -\frac{1}{2T_L}\left\{1 + \sqrt{1 - \frac{4T_L}{T_m}}\right\}$$

$$S_2 = -\frac{1}{2T_L}\left\{1 - \sqrt{1 - \frac{T_L}{4T_m}}\right\}$$

If it is assumed* that $4T_L/T_m << 1$, then

$$\left(1 - \frac{4T_L}{T_m}\right)^{1/2} = 1 - {}^1/_2\left(\frac{4T_L}{T_m}\right) - {}^1/_8\left(\frac{4T_L}{T_m}\right)^2 + \ldots$$

by the binomial expansion. If second order small quantities are neglected, the roots may then be written:

$$S_1 = -\frac{1}{T_L}\left\{1 - \frac{T_L}{T_m}\right\} \cong -\frac{1}{T_L}$$

$$S_2 = -\frac{1}{T_m}$$

Equation 4 then becomes:

$$H_1 = \frac{G_a/K_v}{S(T_m S + 1)(T_L S + 1)} \tag{5}$$

Appendix II. Analysis of the Over-all System

The response function of the over-all system shown in Figure 1 may be written as:

$$\frac{\theta}{e_{in}} = \frac{H_1 H_2}{1 + H_1 H_2 H_3} \tag{6}$$

where

H_1 = response ratio relating the position of the moving coil to the voltage applied to the power amplifier
H_2 = response ratio relating the voltage applied to the power amplifier, to the difference between the first stage input voltage and the position detector output voltage
H_3 = response ratio relating the detector output voltage to the position of the moving coil

---

* In final design $R_T = 500$ ohms, $L = 0.05$ henry, $T_L = L/R = 10^{-4}$, while $T_m = 0.005$. Thus $4T_L/T_m = 0.08$.

As indicated by the bold-face notations, the preceding response functions are written in terms of the Laplace transform variable S. If S is replaced by $j\omega$ the response function becomes the sinusoidal frequency response of the system.[4] With a subscript ω indicating functions of "ω," the frequency response may be written:

$$\frac{\theta_\omega}{e_{in,\omega}}=\frac{H_{1\omega}H_{2\omega}}{1+H_{1.\omega}H_{2.\omega}H_{3.\omega}} \tag{6A}$$

Over the range of frequencies for which

$$H_{1.\omega}H_{2,\omega}H_{3,\omega}>>1 \tag{7}$$

it can be seen that

$$\frac{\theta_\omega}{e_{in.\omega}}=\frac{1}{H_{3.\omega}} \tag{6B}$$

Since $H_{3,\omega}$ relates the output of the detector to the coil position it can, by proper design, be made equal to a constant β over a very wide range of frequencies. The basic over-all sensitivity of the instrument is then given by $1/\beta$. Thus, over the range of frequencies for which condition 7 is satisfied, the system frequency response will be constant.

It is then evident that the prime objective of the over-all system design is to extend this range of frequencies as far as possible. At the same time, care must be taken to insure that the open loop transmission ratio, $H_{1,\omega}$, $H_{2,\omega}$, $H_{3,\omega}$, shall have the proper phase characteristic so that there will be no pronounced resonant peak at the high end of the system frequency response or, even worse, so that the system will not become unstable.[5]

Based on the analysis of the moving system given in Appendix I, the open loop transmission function may be written:

$$A_\omega=H_{1.\omega}H_{2.\omega}H_{3.\omega}=\frac{\beta GG_a/K_v}{j\omega(1+j\omega T_m)(1+j\omega T_L)} \tag{8}$$

where $G=$ the d-c gain of the voltage amplifier.

For the moving coil used in the final design of the system, T_L is sufficiently small so that the factor $(1+j\omega T_L)$ introduces negligible effect over the range 0–300 cycles per second and can therefore be neglected. Equation 8 may then be simplified to:

$$A_\omega=\frac{\beta GG_a/K_v}{j\omega(1+j\omega T_m)} \tag{9}$$

The results of an open loop response test made on the final system prior to equalization are shown in Figure 7. It can be seen that T_m has a value such that $\omega T_m=1$ at a frequency of 35 cycles per second. In order to extend the range of frequencies for which $A_\omega>>1$ and in order to give A_ω the proper phase characteristic, a phase lead equalizer was inserted and the amplifier gain was readjusted. The open loop transmission function then becomes:

$$A_\omega=\frac{T_2}{T_1}\,\frac{G\beta G_a/K_v(1+j\omega T_1)}{j\omega(1+j\omega T_m)(1+j\omega T_2)} \tag{10}$$

where T_1 and T_2 are the time constants of the equalizer.

In the final design $G=4{,}100$, $\beta G_a/K_v=0.330$, $T_m=0.0039$ second, $T_1=T_m=7T_2$. The asymptotic approximation to equation 10 is shown for these values as a dashed line in Figure 7.

The over-all frequency response is obtained by substituting these quantities into equation 6A:

$$\frac{\theta\beta_\omega}{e_{in,\omega}}=\frac{\dfrac{1{,}350}{j\omega(1+j0.00057\omega)}}{1+\dfrac{1{,}350}{j\omega(1+j0.00057\omega)}} \tag{11}$$

The magnitude of the frequency response function is plotted in Figure 8 as a dashed line. The solid line in Figure 8 is the result obtained by actual test. As stated in the body of the article, the disagreement in these curves is due to purely mechanical difficulties in the developmental model which necessitated the addition of a gain reducing equalizer whose effects were not included in the basic analysis presented here.

References

1. AN ELECTRODYNAMIC DIRECT-INKING PEN, **H. B. Shaper.** *Electronics* (New York, N. Y.), March 1946, pages 148–151.

2. A HIGH-GRAIN D-C AMPLIFIER FOR BIO-ELECTRIC RECORDING, **H. Goldberg.** *Electrical Engineering* (*AIEE Transactions*), volume 59, January 1940, pages 60–64.

3. VACUUM TUBE AMPLIFIERS (book), **G. E. Valley, Jr., H. Wallman.** McGraw-Hill Book Company, New York, N. Y., 1948, 743 pages.

4. TRANSIENTS IN LINEAR SYSTEMS (book), **M. F. Gardner, J. L. Barnes.** John Wiley and Sons, Inc., New York, N. Y., 1942, 289 pages.

5. NETWORK ANALYSIS AND FEEDBACK AMPLIFIER DESIGN (book), **H. Bode.** D. Van Nostrand Co., New York, N. Y., 1945, 551 pages.

6. PRINCIPLES OF SERVOMECHANISMS (book), **G. Brown, D. D. Campbell.** John Wiley and Sons, Inc., New York, N. Y., 1948, 400 pages.

7. DIGEST OF PAPERS OF CONFERENCE ON ELECTRONIC INSTRUMENTATION. *Electrical Engineering,* volume 68, number 3, March 1949, pages 246–248.

No Discussion

A High-Sensitivity Ratio Instrument for Industrial Resistance Thermometers

E. W. CLARK
MEMBER AIEE

Synopsis: To make practical the extensive use of resistance thermometers for industrial temperature measurement and control, an instrument having high sensitivity, sturdy construction, and simplicity is needed. To fill this need the ratio instrument described in this paper was developed. Conventional ratio instrument mechanisms proved to be inherently unstable at the high sensitivities required. The mechanism of the new instrument employs a crossed-coil moving system and a novel permanent magnet pole-face geometry resulting in higher sensitivities than have been previously obtained with ratio mechanisms. A resistance thermometer using the new instrument, while sturdy enough to withstand the shocks and vibration encountered in industrial applications, has a sensitivity high enough to make practical scale spans as small as 50 degrees centigrade.

RESISTANCE thermometers have several advantages over other types of temperature measuring equipment[1] especially in the −150- to +150-degree-centigrade range. They are inherently stable and accurate, and the end device[2] may be located several hundred feet from the primary detector without affecting calibration. There are no cold-junction errors requiring compensation, such as are encountered in thermocouple thermometers. Although resistance thermometers have been used for many years in standardizing work in thermometry[3] and for aircraft temperature measurements,[4] they have not been used extensively in industrial applications. The primary reason for this is that instruments which combine high sensitivity and simplicity with sturdy construction have not been available to make practical an industrial resistance thermometer circuit of the type[5] shown in Figure 1.

To fill the need for such an instrument, the high-sensitivity ratio instrument described in this paper was developed. Tentative specifications called for a range of 0–50 degrees centigrade when used with a 10-ohm copper resistance temperature detector, accuracy of ±0.5 per cent of full scale, sturdy construction, a 7-inch scale reasonably linear distribution with means for control initiation, and relative freedom from external influences.

In the circuit shown in Figure 1, the resistance temperature detector and one instrument coil are connected in parallel with a standard resistor and the other instrument coil. A saturating-core transformer is used in the power supply to minimize voltage variations and to keep the current through coil B of the instrument substantially constant. The current through coil A varies inversely with the temperature of the resistance detector. Ambient-temperature influence is minimized by adding a small amount of fixed resistance having a very low temperature co-efficient in series with the instrument coil. A practical minimum value of combined coil and series resistance is approximately 15 ohms. The current through coil A must be kept to a low value in the order of 25 milliamperes, to minimize errors resulting from self-heating of the resistance temperature detector.

Changes in the resistance of lead C will have no effect on the indication, but changes in the resistance of leads A and B will introduce an error that is completely balanced out when I_a and I_b are equal and increases as I_a departs from the value of I_b. This error is minimized by making I_a and I_b equal at a point near center scale. Ordinarily these errors are not appreciable, as leads having very low resistance are used.

The resistance of the resistance temperature detector varies from 9.04 ohms at 0 degrees centigrade to 10.96 ohms at 50 degrees centigrade. Typical circuit conditions for a 0 to 50-degree-centigrade instrument are tabulated in Table I for the center-scale and the end-point temperature conditions. From this tabulation it can be seen that the full-scale deflection must be obtained from a ±4 per cent change in the 25 milliampere center-scale current.

Theory of Ratio Instruments

In order to evaluate the requirements of a ratio instrument for operation on a current differential as low as ±4 per cent, it is necessary to review the theory of ratio instruments. In this discussion, leakage factors and the torques contributed by lead-in spirals are ignored in the interest of simplicity although they must be taken into consideration in actual design. The average value of magnetic flux seen by a coil is assumed to vary inversely with the air gap at the centerline of the coil.

There are two coils on the moving system of a ratio mechanism, so connected that they produce torques in opposing directions. These torques are given by:

$$\begin{aligned} T_a &= -K_a I_a \phi_a \\ T_b &= K_b I_b \phi_b \end{aligned} \qquad (1)$$

where ϕ_a and ϕ_b are the average values of magnetic flux seen by coils A and B, respectively. The minus sign indicates that the torque produced by coil A is in the direction opposite that produced by coil B.

Table I

Temperature, Degrees Centigrade	I_a, Milliamperes	I_b, Milliamperes	I_a/I_b
Low-scale 0	26	25	1.04
Center-scale 25	25	25	1.00
Full-scale 50	24	25	0.96

Paper 52-55, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted October 2, 1951; made available for printing December 10, 1951.

E. W. CLARK is with the General Electric Company, West Lynn, Mass.

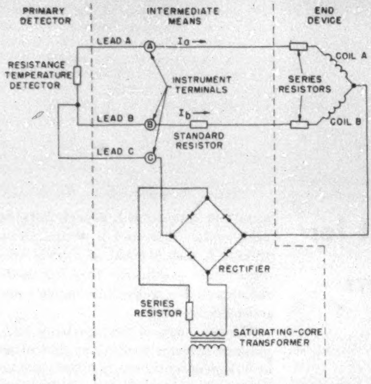

PRIMARY DETECTOR
INTERMEDIATE MEANS
END DEVICE
LEAD A
A
I_a →
COIL A
RESISTANCE TEMPERATURE DETECTOR
INSTRUMENT TERMINALS
SERIES RESISTORS
LEAD B
B
I_b →
COIL B
STANDARD RESISTOR
LEAD C
C
RECTIFIER
SERIES RESISTOR
SATURATING-CORE TRANSFORMER

Figure 1 (left). Resistance thermometer with ratio indicator as end device

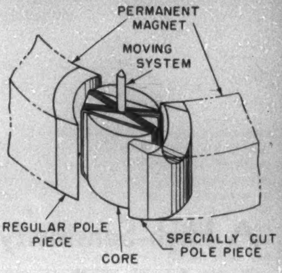

PERMANENT MAGNET
MOVING SYSTEM
REGULAR POLE PIECE
CORE
SPECIALLY CUT POLE PIECE

Figure 4 (right). New cut-pole-face ratio mechanism

The moving system will come to rest when the resultant torque is zero; that is:

$$T_a=-T_b \quad (2)$$

Combining equations 1 and 2:

$$I_a/I_b=(K_b\phi_b)/(K_a\phi_a) \quad (3)$$

One coil may be thought of as producing driving torque while the other coil provides restoring torque. In order to obtain a pointer deflection corresponding to I_a/I_b, these torques must be functions of angular deflection of the moving system. This is accomplished by arranging the geometry of the magnetic elements of the instrument so that the average value of flux seen by each coil will be a function of angular deflection. The relationship of flux values seen by each coil to the ratio of I_a to I_b is given by:

$$\phi_b/\phi_a=(K_aI_a)/(K_bI_b) \quad (4)$$

I_a and I_b may be made equal at center scale by adjusting the number of turns on the coils so that K_a/K_b is equal to the center-scale value of ϕ_b/ϕ_a.

In the following paragraphs, two common types of ratio mechanisms are discussed and it is shown that they are not readily adaptable to the sensitivity requirement set forth above for a 0–50-degree-centigrade instrument.

Off-Center-Core Ratio Mechanism

In the off-center-core ratio mechanism, shown in Figure 2, the air gap seen by coil A increases, while the air gap seen by coil B decreases with angular deflection of the moving system. Since I_a varies inversely with temperature, the driving torque produced by coil A is in a downscale direction while the restoring torque produced by coil B is in an upscale direction. If the air gaps are equal at center scale, K_a may be made equal to K_b so that I_a will equal I_b at this point.

Substituting the conditions set forth previously for a 0 to 50-degree-centigrade instrument into equation 1 and assuming that the gaps are symmetrical, that is:

$$\begin{aligned}&\phi_{a(\max)}=\phi_{b(\max)}=\phi_{\max}\\&\phi_{a(\min)}=\phi_{b(\min)}=\phi_{\min}\end{aligned} \quad (5)$$

the low-scale torques are given by:

$$\begin{aligned}&T_a=-T_b\\&K_a\,1.04\,I_b\,\phi_{\min}=K_bI_b\,\phi_{\max}\end{aligned} \quad (6)$$

and the full-scale torques are given by:

$$\begin{aligned}&T_a=-T_b\\&K_a\,0.96\,I_b\,\phi_{\max}=K_bI_b\,\phi_{\min}\end{aligned} \quad (7)$$

Combining equations 6 and 7 and simplifying:

$$\phi_{\max}/\phi_{\min}=\sqrt{1.04/0.96}=1.042 \quad (8)$$

In terms of air gap dimensions, this means that if the minimum gap is 0.050 inch, a practical value for this type of instrument, the maximum gap would be about 0.052 inch. Such an instrument would be unstable and maintenance of calibration would be extremely difficult as evidenced by the fact that a side shift of

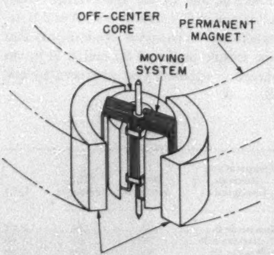

OFF-CENTER CORE
PERMANENT MAGNET
MOVING SYSTEM
POLE PIECES

Figure 2. Two-coil, off-center-core ratio mechanism

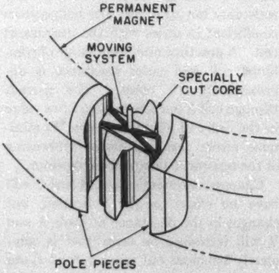

PERMANENT MAGNET
MOVING SYSTEM
SPECIALLY CUT CORE
POLE PIECES

Figure 3. Crossed-coil, eccentric-core ratio mechanism

Table II

Angle of Deflection In Degrees	ϕ_b/ϕ_a	Gap "A" In Inches	Gap "B" In Inches
−18	1.12	0.056	0.050
0	1.10	0.062	0.056
18	1.08	0.067	0.062
36	1.06	0.071	0.067
54	1.04	0.074	0.071
72	1.02	0.075	0.074
90	1.00	0.075	0.075

Table III

Angle of Deflection, Degrees	ϕ_a/ϕ_b	Pole-face Height "A", Inches	"B", Inches
−18	0.94	2.350	2.500
0	0.96	2.256	2.350
18	0.98	2.211	2.256
36	1.00	2.211	2.211
54	1.02	2.255	2.211
72	1.04	2.345	2.255
90	1.06	2.486	2.345

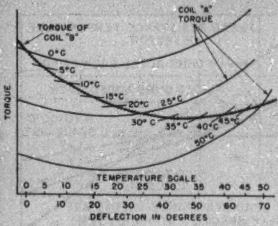


Figure 5. Torque curves of coil A and coil B versus angular deflection of moving system

the core of 0.0001 inch would introduce an error of 5 per cent at center scale.

Crossed-Coil, Eccentric-Core Ratio Mechanism

A crossed-coil, eccentric-core ratio mechanism is shown in Figure 3. In this mechanism the two coils are wound on the same form at an angle of approximately 18 degrees to each other. The relationship between the value of flux seen by the coils and the angular deflection of the moving system is governed by a specially shaped core. Since the two coils operate in the same gap, the ratio ϕ_b/ϕ_a is the ratio of the value of flux at one point in the gap to the value of flux at a point 18 degrees away. The core of this mechanism is usually cut in such a way that the air gap is greatest at one end of the angle of rotation, gradually tapering to its minimum value at the other end. Gap dimensions are calculated on a point-to-point basis from equation 4. Typical air-gap dimensions for a 0 to 50-degree-centigrade instrument having an angle of 18 degrees between the coils and a full-scale deflection of 72 degrees are shown in Table II. Allowance is made for an 18 degree overshoot at each end of the scale. K_a/K_b is made equal to 1.06 so that I_a will equal I_b at center scale.

Although this mechanism is superior to the off-center-core type, it is not readily adaptable to the stated requirements for an instrument having a minimum scale span of 50 degrees centigrade. The effect of side-shift of the core is much smaller than in the off-center-core type, being in the order of 0.2 per cent of full scale for a 0.0001 inch side shift, but the scale distribution of a 7-inch scale is controlled by two tapered cuts approximately .75 inch long and having a maximum depth of 0.0125 inch, making a linear scale very difficult to maintain. A change in the depth of either cut of 0.0001 inch will cause a change in indication of 0.4 per cent.

New Crossed-Coil, Cut-Pole-Face Ratio Mechanism

The fact that it is not practical to increase the sensitivity of either of the conventional mechanisms described above sufficiently to obtain an instrument having a minimum scale span of 50 degrees centigrade led to the development of the new ratio mechanism shown in Figure 4. The moving system is similar to that used in the crossed-coil, eccentric-core mechanism, but the desired relationship between the value of flux seen by the coils and angular deflection is obtained in a novel way. In this mechanism, the core is concentric with the pole faces providing a uniform air gap, but one pole face is cut in such a way as to reduce the effective height of the pole. Neglecting the effect of leakage flux distribution, the average value of flux seen by the coil varies with the effective height of the pole at the center line of the coil. The ratio action in this mechanism is accomplished by a variation in the height of the pole face which is in the order of 1.25 inches while in the previously described mechanism ratio action is accomplished by variation in the air gap which is in the order of 0.050 inch. Obviously, the dimensions of the cut in the new mechanism are much less critical and easier to maintain.

Equations 1 through 4 hold for this mechanism and the required depth of the pole-face cut may be calculated on a point-to-point basis from equation 4. Typical pole-face dimensions are given in Table III for a 0–50-degree-centigrade instrument having an 18 degree angle between coils, a full-scale deflection of 72 degrees, and a maximum effective pole-face height of 2.5 inches (the sum of left and right 1.25 inch sides.) Allowance is made for an 18 degree overshoot at each end of the scale. This sample calculation shows that the sensitivity requirements of a 0–50-degree-centigrade instrument can be met quite easily with the new crossed-coil, cut-pole-face ratio mechanism. The pole-face cut, which in the example has a depth of 0.289 inch, is obviously much easier to maintain than the much smaller cut required in the mechanism which was described previously. In this mechanism, the effect of side shift of the core is negligible because the air gap is concentric.

In Figure 5, a family of curves representing the torque of coil A at various temperatures and the constant-current torque of coil B are plotted against angular deflection of the moving system. The point at which each torque-curve of coil A intersects the torque-curve of coil B corresponds to the angular deflection at that temperature. These curves show that the scale distribution is very nearly linear.

Table IV

	% of Full Scale
Rated Accuracy 120 volts, 60 cycles per second, 25 degrees centigrade	±0.5
Influences	
Voltage, ±10% of rated voltage	±0.1
Frequency, ±1 cycle from 60 cycles per second	±0.4
External field, 5 oersteds	less than 0.1
Response Time*	2.5 seconds

*The response time is a function of the total resistance in the circuit and will be longer for instruments having larger scale spans. The response time of a 20 to 140-degree-centigrade instrument is 4 seconds.

Application to Temperature Measurement

Figure 6 shows a complete temperature meter incorporating the new crossed-coil, cut-pole-face mechanism. Vertical pivot construction, a very high strength Alnico V magnet and low-torque current-lead-in spirals combine to produce high torque and to minimize external influences.

Figure 6 (left). Ratio-type temperature meter incorporating new mechanism

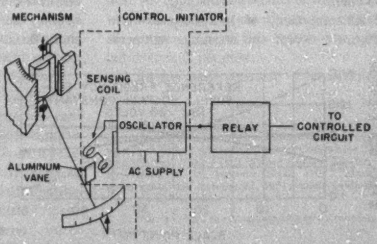


Figure 7 (right). Control initiator

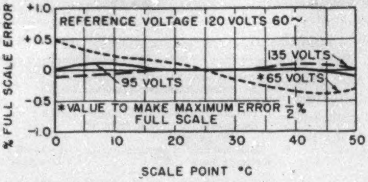


Figure 8 (left). Voltage influence

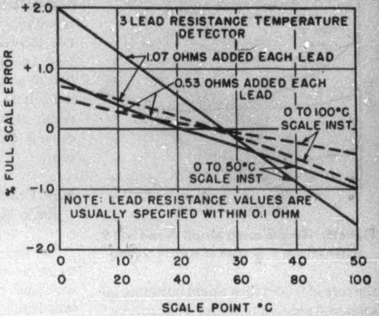


Figure 10 (right). Influence of changes in lead resistance

A commonly accepted criterion for comparing electric instruments is the torque-to-weight ratio and for vertical pivot instruments the ratio of torque to the one-and-one-half power of the weight.[6] For this instrument:

$T/W = 1.02$
$T/W^{1.5} = 0.73$

where torque T is in millimeter grams, and weight W is in grams.

This compares very favorably with conventional ratio instruments. A typical crossed-coil, eccentric core, ratio temperature meter having a full scale span of 100 degree centigrade has a factor of merit of:

$T/W = 0.48$
$T/W^{1.5} = 0.32$

Control Initiation

Control initiation[7] is provided by a small aluminum vane which is mounted on the pointer arm and two sensing coils which are so arranged that the aluminum vane may enter the space between them, see Figure 7. When the vane enters the space between the coils, their effective inductance is reduced. The position of the sensing coils may be adjusted so that control initiation is provided at any temperature within the range of the instrument. Usually the sensing coils are connected in an electronic oscillator circuit which is adjusted so that it will oscillate only when the vane is in the space between the sensing coils. A relay in the plate circuit is adjusted so that it is closed when the circuit is not oscillating. When the vane enters the space between the coils, the circuit oscillates and the accompanying reduction in plate current causes the relay to open.

All components of the system shown in Figure 1 except the resistance temperature detector, as well as the automatic control system are contained in the instrument case.

Performance Characteristics

The performance characteristics of the instrument are shown in Table IV.

The performance characteristics of a 0 to 50-degree-centigrade instrument are given in more detail in Figures 8 through 10. The 0–50-degree-centigrade range was chosen for preparation of these data because it is representative of the most severe conditions imposed on the instrument. The influence of lead resistance variation and ambient temperature are much smaller in instruments having larger scale spans. This is shown in Figure 10 where the influence of lead resistance variations on an instrument having a 50-degree scale span is compared with that on an instrument having a 100-degree scale span.

Conclusions

The new crossed-coil, cut-pole-face ratio mechanism makes practical the extensive use of resistance thermometers for industrial temperature measurement and control. The new cut-pole-face construction, combined with a crossed-coil moving element, a very high strength Alnico V magnet, low-torque spirals and a reliable means of control initiation, results in a resistance thermometer having the following characteristics:

1. High sensitivity (scale spans as small as 50 degrees centigrade).
2. High accuracy (±0.5 per cent of full scale).
3. Sturdy construction.
4. Legibility (7 inch linear scale).
5. Simplicity (manufacturing and maintenance problems less critical than in prior instruments).
6. Versatility (may be changed from indicator to controller by adding electronic relay).

The mechanism is not limited to temperature measurement, but is adaptable to any application where the ratio of two currents is to be measured over ranges as small as ±4 per cent of center scale current in one circuit.

References

1. Electric Controllers for Laboratory Furnaces, M. H. Roberts. *Electronic Engineering* (London, England), volume 23, February 1951, number 276, pages 51–54.

2. Functional Analysis of Measurements, I. F. Kinnard. *Electrical Engineering (AIEE Transactions)*, volume 66, January 1947, pages 11–15.

3. The International Temperature Scale of 1948, H. F. Stimson. Research Paper *RP-1962*, volume 42, March 1949, U. S. Department of Commerce, National Bureau of Standards.

4. A Unique Moving-Magnet Ratio Instrument, F. R. Sias, D. B. Fisk. *Electrical Engineering (AIEE Transactions)*, volume 63, September 1944, pages 634–36.

5. Advantages of Resistance Thermometers Paul G. Weiller. *Electrical Manufacturing* (New York, N. Y.), volume 44, number 3, September 1949, pages 104–07, 204.

6. Mechanical Factor of Merit with Respect to Electric Instruments, J. H. Gross. *General Electric Review* (Schenectady, N. Y.), April 1933, pages 188–99.

7. Pyrometer Millivoltmeter and Automatic Controller, E. W. Clark, J. K. Sidebottom. *Mechanical Engineering* (New York, N. Y.), November 1951, page 934.

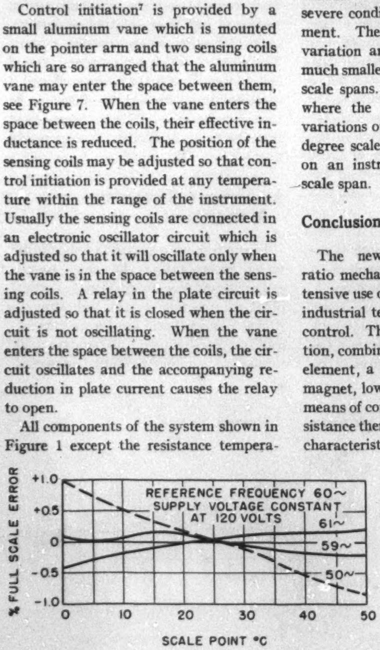


Figure 9. Frequency influence

No Discussion

Methods and Apparatus Employed in the Testing of Single-Phase A-C Watthour Meters

F. C. HOLTZ
FELLOW AIEE

DURING the past decade much attention has been directed to improving the accuracy and speed of meter testing on the part of manufacturers and users of a-c watthour meters. Numerous arrangements are in general use which might well be publicized for the benefit of those who feel the need for improved methods and equipment.

Much of the present paper is historical and is primarily intended to indicate a few of the factors which should be considered in the evaluation of testing methods. It describes one form of apparatus which has for its object the reduction of personal errors and which makes available a permanent record of individual tests. It is the author's hope that it will stimulate interest in the general subject and encourage the publication of new material.

The object of this paper is to review briefly the various methods which are available for the testing of single-phase a-c watthour meters and to indicate a general trend which is aimed at securing greater accuracy in the final results. The problems encountered in the precise testing of multielement meters are of a more complicated nature and are not made a part of this paper.

The testing of a-c watthour meters presents a special problem to the manufacturer and user of these devices since there is not available a fundamental standard of reference such as is possible for the testing of indicating instruments to be used later on alternating current. In order to insure the desired order of accuracy in the finished product it is customary for the meter laboratory to have access to and maintain a standards laboratory which may, at all times, duplicate to the required degree of accuracy the fundamental units such as the volt, ohm, and second.

The units which are used in practical measurement are the absolute units defined in terms of concrete standards adopted by an Act of Congress in 1949.[1]

One of the responsibilities of the National Bureau of Standards, United States Department of Commerce, is to maintain such standards of reference and to test and issue certificates of test, evaluating comparison standards in terms of these absolute units.

The best available standard of electromotive force is perhaps the cadmium cell, usually referred to as the Weston Standard Cell. Such cells purchased from reliable sources and verified by the Bureau of Standards can usually be relied upon to be correct to at least 1 part in 10,000. The Bureau will likewise certify standard resistances made of manganin to an accuracy of at least one part in 10,000.

The testing laboratory may therefore avail itself of at least two of these fundamental units. A third, the unit of current, can be derived from these two by means of auxiliary equipment such as potentiometers, volt boxes, et cetera.

Such equipment cannot be used for the direct measurement of alternating current and it is therefore necessary to make use of instruments of a type which may be calibrated with direct current and then used with equal facility on alternating current. The most important of these are the so-called "dynamometer" wattmeters, whether of the "laboratory standard" type or the conventional portable type extensively used in a-c measurements.

In an effort to improve upon the accuracy of these devices for the measurement of alternating current the Bureau of Standards has, over the past few years, developed for its own use special dynamometer instruments capable of a very high order of accuracy. The so-called "glass wattmeter"[2] used there provides a means for the measurement of a-c watts to an accuracy of better than 0.02 per cent.

Similar developments have occurred in England[3] where the A-C/D-C Comparator of "Shotter-Hawkes" is manufactured commercially by Elliott Brothers for several of the Electricity Board Central Laboratories. This instrument, illustrated in Figure 1, comprises a carefully made dynamometer on the moving system of which is also a d'Arsonval d-c ammeter of limited current capacity. In this arrangement the apparatus is first calibrated by connecting the current elements in series and checking with direct current. Later, the dynamometer elements are connected into the a-c circuit to be measured and a balance is obtained by passing through the d-c instrument a direct current which is evaluated by means of a potentiometer set up using standards of electromotive force and resistance. One such instrument, recently brought to this country, was checked first at National Physical Laboratories in England and later by the National Bureau of Standards on arrival in the United States. Not only were the comparative results of the checks very close, but the maximum error over the entire range of this instrument was not over 0.02 per cent.

A further basic constant of importance

Paper 52-56, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted October 15, 1951; made available for printing December 10, 1951.

F. C. HOLTZ is with the Sangamo Electric Company, Springfield, Ill.

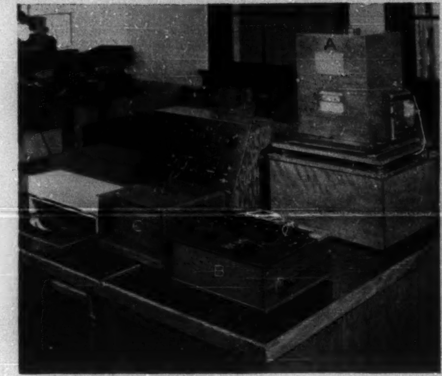


Figure 1. Shotter-Hawkes a-c/d-c comparator

(A)—comparator; (B)—control box; (C)—standard transformer

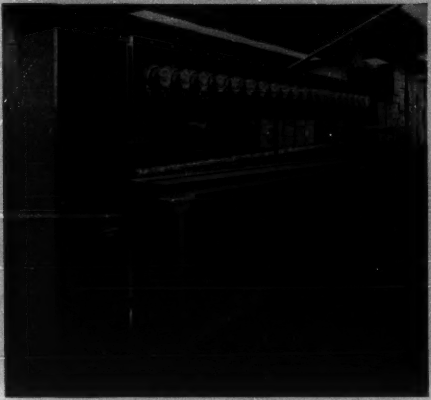


Figure 2. Typical photoelectric test board used by one large public utility laboratory

to the laboratory in the testing of watthour meters is the element of time. This is a unit which is made available to a very high order of accuracy. The Bureau now sends out time signals over station WWV to an accuracy of better than one part in 200,000. Standard laboratory clocks and timing devices can therefore be checked at frequent intervals and their accuracy established to a degree beyond any ordinary commercial requirements.

In his classical paper on The Accuracy of Commercial Electrical Measurements[4] Dr. Herbert Brooks of the National Bureau of Standards has set forth with great clarity and detail the need for accuracy in commercial electrical measurements and the means by which it may be attained. His paper deals almost exclusively with indicating instruments where all observations are of an instantaneous nature. In the testing of watthour meters which are integrating devices such instruments are an essential part of the method of determining their accuracy.

In this process, indicating instruments are used in combination with various forms of time measuring devices and through a variety of methods, the accuracy of the watthour meter is arrived at. It may be of interest to review a number of these methods and point out the sources of error involved. In the present author's opinion, they may conveniently be divided into three categories, as follows:

I. Direct methods using indicating instruments and timing devices.

II. Indirect or comparison methods using rotating standards.

III. Methods using indicating instruments in which the measurement of time is not directly involved.

Each of these methods is susceptible of many variations. It is to be noted that the first and third category lie nearest the primary standards, for here we are using reference instruments which through the medium of a potentiometer have been checked against primary standards of electromotive force and resistance, and then with adequate assurance, used on alternating current.

The second category is merely an extension of the first in which a watthour meter, after having been calibrated by the first method, is used as a reference standard in the second. This has lead to a reasonably successful attempt to use one or more watthour meters as standards of the watthour.[5] The indirect or comparison method using rotating standards is capable of many variations. Included in this group are stroboscopic methods and what is considered to be a new arrangement described more in detail by the author. As will be shown later, a third category is possible which is, in reality a combination of I and II.

I. The Direct Method

The application of the direct method recognizes the fact that the watthour meter is an integrating instrument and that its per cent registration in terms of the wattmeter readings and time is given by the following equation:

$$P=\frac{100K_sN}{WT} \qquad (1)$$

where

K_s = watt-second constant or watt-seconds per disk revolution of the meter under test

N = number of meter disk revolutions

T = time in seconds for N revolution of the disk

W = load in watts as determined from wattmeter reading

In the simplest arrangement the load is held reasonably steady by manual control and by means of a stop watch, the time required for the watthour meter to make an integral number of revolutions is noted. The error involved in such a process depends upon a number of factors not the least of which are those of a personal nature. There is, for example, the personal error involved in holding the load steady and if not steady, the ability to hold to an average value over the interval during which the time readings are being taken. Additionally, there is the error on the part of the observer in starting and stopping the clock at the precise instant that the mark on the disk is in line with the reference point. These errors may be reduced by taking a sufficiently large number of disk revolutions during any individual test. Within recent years the application of this method has been greatly improved through the use of constant voltage transformers which permit holding the load more steady or as a further elaboration through the application of electronic apparatus which hold a given load automatically.

Within the past few months electronic circuits have been designed and constructed capable of maintaining a-c voltage constant to 0.01 volt under input variations customarily encountered in service.[8]

A further elaboration comprises in the application of photoelectric cells which automatically start and stop the timing device and therefore substantially eliminate the element of personal error in the determination of the time for completing a given number of disk revolutions. It will therefore be appreciated that through the application of these devices it is possible to determine the accuracy of a watthour meter to a high degree and make it dependent, in a large measure, on the accuracy of the comparison instrument. Since this method provides for the continuous rotation of the watthour meter it should be considered the most satisfactory.

Another procedure associated with the first category makes use of an accurate seconds pendulum which, through a system of relays, starts the meter from some zero position at one instant, and after a given integral number of seconds have elapsed, disconnects it from the circuit. In this event, the precise integral and fractional parts of meter disk revolutions must be determined. While an acceptable procedure, it is not as accurate for reasons which follow.

Two kinds of errors associated with this method may be attributed to the meter under test. They are:

1. Errors due to the starting and stopping of the disk.
2. Inability to maintain the disk stationary both before and after each run.

The accepted practice in this type of test is to keep the potential circuit of the meter excited and start and stop the meter by closing and opening the current circuit. Since the potential circuit of the meter is highly inductive, this practice avoids the possible effect of transients which might otherwise be present during a starting period.

Under group 1 above we may consider the effect of the inertia of the disk which must be overcome at the commencement of rotation and which is active during periods of deceleration.

It can be shown that if all the friction of the moving system has been compensated for by the light load adjustment, the loss in rotation during the period of starting is exactly compensated for by a gain in rotation after the current circuit of the meter has been broken.

Under group 2 the effect which must be carefully watched is any tendency for the disk to creep away from its set position before starting due to the effect of the anticreep slots or other causes which may be active during the interval before the meter is started. The same precautions should be taken with reference to the so-called "rest" position of the disk after the circuit has been opened.

The presence of any vibration will usually contribute materially to this difficulty. These effects may be slightly variable but in any event, they become more serious when, as on light load, fewer revolutions are recorded. It is highly advisable to select a time interval T such that for correct registration the meter disk should make an integral number of revolutions. The recording of partial revolutions of the disk brings into play variations in disk speeds which may introduce appreciable error. These will be discussed in more detail under stroboscopic testing.

The degree of accuracy required in the final result will dictate the refinements by way of technique and procedure which should be followed. It is useless, however, to carry this beyond the point where the reliability of the meter under test may be in question.

When the meter under test is to be used as a "master meter" in some of the comparison methods to be discussed later it should be selected for reliability and the tests should be made by the direct use of the dynamometer type a-c/d-c comparators previously described. For ordinary use in checking commercial watthour meters the standard indicating dynamometer-type wattmeters are sufficiently accurate for use as transfer instruments.

As an example of the probable error involved in the final test of commercial watthour meters as practiced by the average smaller central station laboratory, the following is representative:

1. Primary standard ±0.025 per cent.
2. Reference standard ±0.05 per cent.
3. Working standard ±0.20 per cent.
4. Application of working standard in calibration of watthour meter ±0.20 per cent.

The expected probable error (rms of above errors) ±0.3 per cent.

The importance of recognizing and reducing all these errors must be stressed since the application of the above two methods is quite general in arriving at a unit of watthours or watt-seconds for laboratory or field testing by means of the second group called here indirect or comparison methods. This might well be considered "minimum" requirements to be reduced where necessary by the elimination of personal errors through the utilization of automatic equipment but not to be extended beyond the reliable performance of the meter under test.

II. Indirect or Comparison Methods

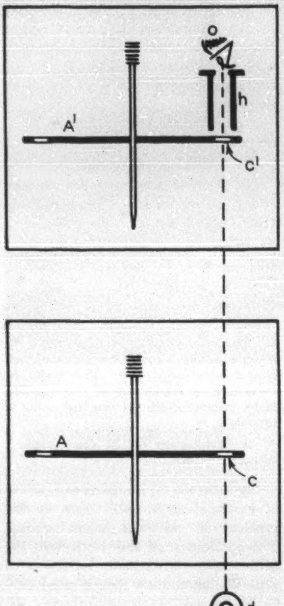
The methods coming under category II may be divided into a number of subgroups as follows:

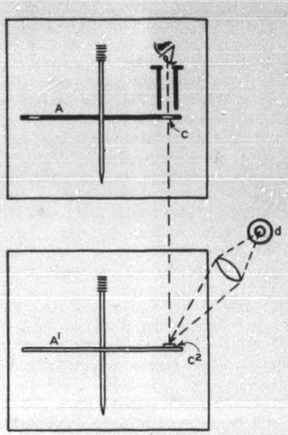
1. Directly comparing revolutions of the meter under test with those of a rotating standard under the same load.
2. Stroboscopic methods in which the speeds of the meters are directly compared.
3. The watthour comparator in which the relative rotations are automatically recorded on a strip chart. This might be considered a refinement of (1) capable, under proper conditions, of a high order of precision.
4. A modification of (3) above using a cathode-ray tube.

The subgroup (1) is capable of a variety of modifications to meet some specific problem in watthour meter testing. The application of this method to the field testing of watthour meters is so generally understood that it need only be mentioned. In some instances the load is directly applied through the use of a resistance or load box whereas in others phantom loading is resorted to in order to decrease the weight of the equipment as well as the energy consumed during the test.

The likelihood of error is considerably increased due to unsatisfactory operating conditions such as rapid or extreme changes in temperature to which the rotating standards are subject, the effect of handling and in some instances the influence of an unfavorable location of the meter under test. It must be said however that despite these apparent disadvantages the results obtained are well within the requirements of all regulatory bodies. In some instances the difficulties encountered in field testing have led to the practice of returning to the shop all meters which require periodic test and inspection.

For the shop testing of meters it is customary to gang test a number of meters of the same rating simultaneously against a reference standard. In this event the current coils of all the meters under test, including that of the rotating standard are connected in series with a resistance and the low voltage winding of a transformer capable of supplying full load current. For separately exciting each of the potential coils of the meter





Figures 3 (left) and 4 (above). Stroboscopic testing by direct and reflected light

under test a special transformer is employed having the required number of secondary coils of like characteristics, each being connected to a potential coil of a watthour meter. This insures like voltage, current, and power factor being applied to all the meters involved in the test.

Preceding the test, the register of the rotating standard is set on zero and the spot on the disk of each meter is aligned with some reference point, usually one edge of a permanent magnet. With the potential circuits all previously excited, the current circuit is now closed until the rotating standard has made a designated number of revolutions at which point the current circuit is opened. The accuracy of each meter is then determined from the fraction of a revolution by which the reference mark on each meter disk deviates from its zero position. To facilitate such measurement the periphery of each disk is usually divided into 100 or more equally spaced divisions. The difficulty in stopping the disk of the rotating standard at the precise point has introduced the use of photoelectric means which permits the rotating standard to run continuously and by means of relays provides for starting the meters under test corresponding to some selected position of the rotating standard disk and stopping them after an integral number of revolutions of the standard has taken place. In any of these arrangements either the current or potential circuits may be excited continuously and the other interrupted. Because of the low power requirements of the potential circuits, this one is usually interrupted.

For light load testing the current coil of the rotating standard is usually fed from a 1/10 current transformer or a current range provided such that the standard will operate at full load speed while the meters under test are operating at 1/10 load. In order to provide for improved accuracy at light load the equipment is so adjusted that the meter under test makes at least two revolutions.

A slight modification of the arrangement is made when one meter is tested at a time. In that event the photoelectric cell and light source is applied to the meter under test and by the same system of relay counting, the rotating standard is automatically started and stopped.[7,8] Photoelectric means are now very generally applied to the gang testing of watthour meters, since it not only eliminates personal errors but it permits the operator to perform other minor duties while the meters are under rotation.

Figure 2 is a view of a modern public utility laboratory test bench where 20 meters are simultaneously tested and adjusted by this method. It is appreciated that these are only typical examples of a general method which is capable of many variations.

The stroboscopic method outlined under (2) is by no means new. The stroboscopic principle as applied to rotating elements may be explained as follows:

If the uniformly rotating meter disk, provided with a series of equally spaced peripheral markings is viewed through a shutter or by means of light pulses of short duration and of constant frequency the stroboscopic effect will be observed. If these pulses come at the instant when the position of the markings exactly overlap those previously observed, the pattern will appear stationary. If, however, their positions are just in advance or behind those previously noted, the pattern will appear to be moving in the direction of rotation or opposite to it respectively, the rate depending upon the amount of advance or retardation observed between pulses. The pulses must, in addition, be sufficiently rapid and short in duration such that with the persistence of vision, they appear as a uniformly rotating pattern. In general, the lower limit of pulses for satisfactory viewing is about 16 per second below which the sense of continuity disappears. While there is no sensible upper limit, the sense of rotation disappears when the apparent speed of the images reaches an upper limit in the same way that the eye cannot under some conditions discern the spokes of a rapidly rotating wheel. The latter is of no consequence in the present method whereas the former limits its use in light load testing.

One of the earliest applications of stroboscopic testing was patented and used by Otto Blathy of Austria Hungary in 1916. This comprises, as illustrated in Figure 3, in placing the master meter directly above the meter under test with their disks rotating substantially about a common axis. Each disk is provided with an equal number of uniformly spaced openings. A tubular aperture is mounted just above the disk of the master meter and in line with the equally spaced openings such that the latter may act as a shutter and permit intermittent viewing of the slots or openings in the disk of the meter under test.

Both meters, which are assumed to be of the same capacity, are connected to the same load such that their current coils are in series and potential coils in parallel. A light source is placed below the lower disk on the axis of the viewing tube such that it is exposed to view each time one of the openings in the lower disk lines up with the axis of the tube. Assuming the rotating standard to be correctly calibrated and the load adjusted to the desired value, the meter under test is adjusted until the light pulses become stationary. With this method it was not easy to determine whether a meter was fast or slow.

An alternative and simpler method consisted in marking the disk with an equal number of uniformly spaced radial lines which were illuminated by a source of light. In this arrangement the stroboscopic effect was observed and it was possible to determine whether the meter was fast or slow, Figure 4. The precise amount becomes a matter of determining the relative motion of the image with respect to the actual motion of the disk. In general, this requires the introduction of the element of time since the speed is not directly observable. If one knows the speed of the standard and the number of peripheral openings, the number of pulses per second can readily be determined. If we then determine the speed with which the images appear to be passing a given index line per second the accuracy of the meter under test can be determined. For example, if a master meter making $16^{2}/_{3}$ revolutions per minute is provided with 180 openings the frequency of the flashes will be $16^{2}/_{3} \times 180/60 = 50$ cycles per second. If the image appears to be moving forward with a speed of, say 1/4 division per second the meter under test would be 0.5 per cent fast. This is an amount which is readily observable. It is to be noted that this is the same error which might have been observed using a start-and-stop method with the same meters making the same number of revolutions in the same period of time. Hence it may be said that the stroboscopic method is incapable of any greater accuracy nor is it less time-consuming than the start-and-stop method. Its great advantage lies in the fact that it requires so little time to observe the state of calibration of a meter and hence make adjustments while the meters are under rotation. Another important advantage lies in the fact that both meters are rotating continuously and starting and stopping errors are eliminated. These are features which, alone, would recommend the method wherever conditions permit. It has the disadvantage that any nonuniformity in the disk gives rise to a nonuniform rotation and if an attempt is made to use it on light load the lines must be kept under constant observation and the algebraic sum of the images passing a given point must be used. This happens when the meter is so near calibration that the nonuniform motion of the image is first in one direction and then in the other. Despite these difficulties, however, this method has been successfully applied to light load testing by using a somewhat more elaborate arrangement.

In such cases, the periphery of the disk of the meter under test is provided with, say 720 equally spaced teeth. The standard is usually operated at 10 times the speed of the meter being tested and is arranged to intercept the light from a rather powerful source which, by means of lenses, provides a parallel beam which grazes the edge of the disk of the meter under test. By means of a projecting lens the magnified image of the peripheral edge of the disk is projected on a ground glass screen conveniently located with respect to the operator. With the same equipment and through the choice of appropriate current transformers, the meter may be calibrated at a variety of loads.

In a further modification[8] of the method the rotating standard intercepts the light falling upon a photoelectric cell. By means of a suitable amplifier, power is fed to a small synchronous motor, the shaft of which carries a disk having equally spaced peripheral slots. These again act as shutters and intercept the light falling on the disk of the meter under test. Through a suitable selection of slots, the light is pulsed in accordance with the required rates already indicated. This provides for somewhat greater flexibility since the rotating standard can then be used for simultaneously adjusting a number of meters as in gang testing.

More recently a very ingenious arrangement was described before the Institution of Electrical Engineers of London.[9] In this arrangement an optical system is employed which provides for keeping the markings on the disk of the meter being tested under continuous observation. Its advantages appear to reside in the removal of the so-called stroboscopic effect thereby rendering it useful for light-load testing but it is otherwise troubled by factors inherent to the stroboscopic method.[8]

The Watthour Meter Comparator

Another device to which the author has recently directed his attention appears to possess a number of advantages over many of the arrangements already described. The basic element is a recorder previously employed in determining the rate of timepieces in which the essential information is automatically and continuously printed on a tape. This device as applied to the testing of watthour meters may best be described by reference to Figures 5, 6, and 7.

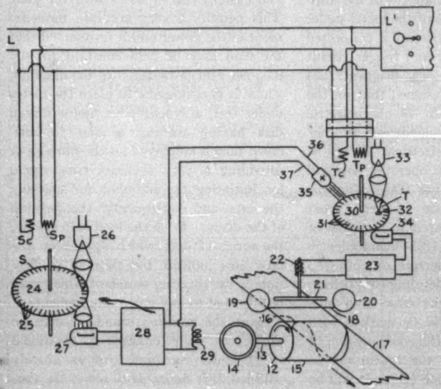
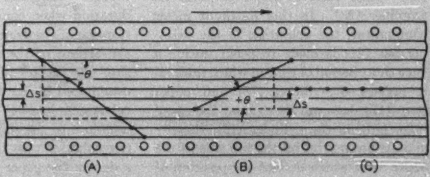


Figure 5 (left). Line schematic diagram illustrating use of the watthour meter comparator

Figure 6 (below). Section of tape illustrating various types of records



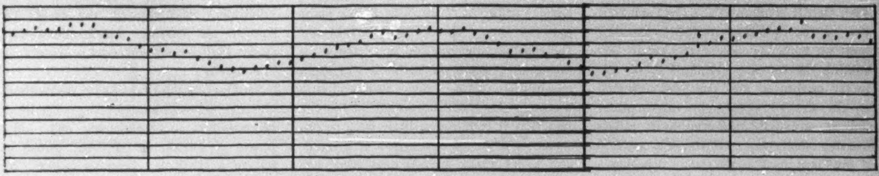


Figure 7. Section of tape illustrating variations in speed of rotating standard

Referring to Figure 5, the main recording element comprises a cylinder (12) mounted on, or driven by the shaft (13) of a synchronous motor indicated at (14). In the illustrated arrangement the cylinder (12) is operated at the same speed as the synchronous motor (14) but it may be advantageous to provide a selective speed drive for the cylinder (12), so that the cylinder can be operated at a satisfactory speed for the testing of meters of different load capacities, such as full load, light load, et cetera. The surface of the cylinder is smooth except for a spiral or helical ridge (16), of substantially triangular cross-section the function of which will now be described.

Moving tangentially to, but out of contact with cylinder (12), is a paper tape (17) which is driven by means of gearing to the synchronous motor shaft (13).

Situated adjacent to the paper strip (17), and parallel to the axis of the cylinder is a conventional typewriter ribbon (18), with its spools (19) and (20) driven by suitable means for the purpose of presenting a fresh ribbon surface to the paper strip (17) as often as necessary. A convenient arrangement for driving the ribbon is by a reduction gear drive connecting to the shaft (13) of the synchronous motor (14).

Immediately above and in close proximity to the typewriter ribbon is a printing bar (21), whose flat surface lies parallel to the axis of the cylinder (12). In practice the printing bar (21), ribbon (18), and paper strip, are almost in contact with each other along a line parallel to the shaft (13) and axis of the cylinder (12). The printing bar (21) is arranged to be actuated by a suitable solenoid (22) which receives impulses from an amplifier (23). In this watthour meter comparator, the operation of the printing bar (21) is made to occur during a very short period of time in which interval it presses the ribbon (18) and paper strip (17) downwardly against the spiral ridge (16) on the cylinder (12), thereby making an imprint on the top side of the paper strip which defines the position of the spiral ridge (16) at the particular instant when the print impulse occurred.

The combination of the rotating standard with its photoelectric pickup and amplifier driving the synchronous motor is similar to an arrangement described elsewhere.[8]

The disk (24) of the rotating standard is provided with a series of equally spaced peripheral slots (25) which serve to alternately permit the light from a source (26) to impinge upon a photoelectric cell (27). By means of the amplifier (28) the field coil (29) of a synchronous motor (14) is supplied with an alternating current having a frequency which is directly proportional to the speed of the meter disk (24).

The disk (30) of the meter under test is provided with the conventional stroboscopic markings (31) generally the same in number as the slots (25) in the peripheral edge of the disk of the rotating standard. A stroboscopic light source (35) connected to the amplifier (28) enables a preliminary adjustment to be made in the speed of the meter under test, primarily for the purpose of bringing its accuracy within the range of the comparator, which in the present device is approximately 1 per cent.

A conventional photoelectric cell and light source is associated with the anti-creep slots in the disk of the meter under test. The photoelectric cell is connected to a special amplifier and trigger circuit indicated by the block diagram (23) the latter being so designed that on the first entry of light into the photoelectric cell (34), a charged condenser or other suitable source of electrical energy is discharged through the operating solenoid (22) of the printing bar (21). The discharges therefrom are of very short duration when compared to the time of rotation of the drum (12). Consequently the arrival of the aperture or anticreep slot (32) at its predetermined position established by the light source and optical system (33-34) will be marked with reference to the simultaneous position of the cylinder (12) by the appearance of a dot on the paper tape (17). It will be noted that with a uniform rotation of the cylinder (12) in a clockwise direction, as indicated by the arrow, the contact position between the spiral (16) and the printing bar (20) will progress uniformly from left to right across the cylinder, and may occur at any point along a line which is the axial length of one complete revolution of the spiral (16).

The spiral therefore serves indirectly as a means for measuring fractional parts of a drum revolution. This, in combination with the synchronous motor and slots in the disk of the rotating standard serves to greatly magnify its rotational speed and enables its position to be more accurately determined.

It may be of interest to outline briefly the procedure followed in the calibration and use of the comparator in what might be termed precision testing of a watthour meter.

CALIBRATION

In the calibration of the standard, the arrangement shown in Figure 5 is modified as follows:

Using a power supply capable of very close and automatic regulation, the rotating standard is connected in circuit with an a-c/d-c comparator such that the potential coils are in parallel across the supply and current coil in series with the load. This permits a very accurate measurement of the power and it is assumed that the load may be held constant to, say 0.05 per cent or better. In the diagram, where it is customary to place the meter under test, a synchronous motor driven disk having openings similar to anti-creep slots is employed for the purpose of providing a very accurate time signal for initiating the impulses for marking the tape and consequently the position of the drum. With the load held steady, the series of dots which appear on the tape now indicate the state of calibration of the rotating standard which may be adjusted to any given requirements.

Since the rotating standard may be subject to slight changes in rotational speed under constant load as already outlined, it is desirable to select the timing impulses such that under conditions of accurate calibration the former always makes an integral number of revolutions per impulse. After calibrating the rotating standard the timing device is removed and a watthour meter may then be substituted and checked or calibrated by means of the rotating standard as described.

Assuming that the meter is to be calibrated, the load is first adjusted and maintained at approximately the same value as used in the initial calibration of the standard. The meter under test is then brought within the range of the comparator by adjusting it stroboscopically as previously described. With the comparator circuits energized, a series of dots will now appear on the tape according to one of the three general patterns illustrated in Figure 6.

As shown in the Appendix, the spacing of the dots as measured across the tape corresponding to successive meter disk revolutions or the angle which they make with the direction of motion of the tape will be a measure of the per cent error of the meter under test with reference to the standard.

This is given by the equation

$$E=-\frac{100\alpha}{n_d+\alpha}$$

If α is small, this becomes

$$E=-\frac{100\alpha}{n_d}$$

These equations are based upon the assumption that the meter under test has the same constant as the rotating standard. The example illustrated in the Appendix provides for 180 slots in the disk of the rotating standard and for a 6 pole synchronous motor directly coupled to the drum. With this assumption the drum is rotating 60 times faster than the standard. With a drum diameter of 1 inch and a spiral pitch of 2 inches, the error of the meter under test in terms of measurements taken from the tape is $E=-0.833\Delta S$ where ΔS is the distance in inches, between dots as measured across the tape corresponding to one meter disk revolution. For example, if ΔS equals 0.12 inch, the error is 0.1 per cent which can easily be measured. When testing a meter at 10 per cent load it is customary to maintain full load speed of the standard. This is accomplished by introducing a current transformer into the circuit such that with full load on the standard exactly 10 per cent current flows in the current coils of the meter under test. It is to be noted that under this condition the standard makes ten revolutions to one of the meter under test. If the equipment remains unaltered the drum of the recorder will now make 600 revolutions between successive meter disk revolutions if in calibration. Under this condition the error in terms of ΔS becomes

$$E=-0.0833\Delta S$$

If ΔS has a value of 0.12 inch the error is 0.01 per cent and the device becomes too sensitive for any practical purpose. It is therefore customary to use a second recorder for light load testing in which the motor is geared to the drum in a substantial ratio.

For example, using a ratio of, say 45/4 and a 16-pole synchronous motor, the drum revolves only twice as fast as the master meter and

$$E=-\frac{100\Delta S}{40}=-2.5\Delta S$$

Under these conditions 0.04 inch spacing between dots for one revolution of the meter disk is 0.1 per cent. It is to be noted that if the dots proceed in one direction with respect to the direction of motion of the tape, the error will be positive and in the other, negative. This is illustrated in Figure 6 in which (A) illustrates the nature of the record when the meter under test is fast, (B) when slow, and (C) correct with reference to the master meter.

Since the tape is driven by the synchronous motor, the slope of the dots will also indicate directly the extent of error. For example, in the present design the constants were so chosen as to make $\tan\theta=-10E$; hence for an error of 0.1 per cent the slope of the line is 45 degrees in one direction, if the meter is fast, and in the other if it is slow.

The condition of the meter under test may be quickly ascertained by means of a small protractor having an index line which may be lined up with the succession of dots on the tape and the error read from suitably calibrated scale engraved on its periphery with reference to a fixed index on the frame carrying the protractor.

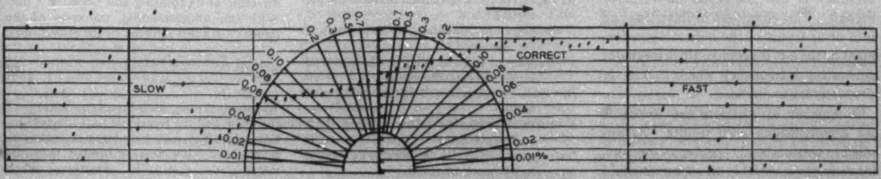
The comparator in its present form is a very useful instrument as a laboratory device for determining small variations in the meter under test. As examples, are the following:

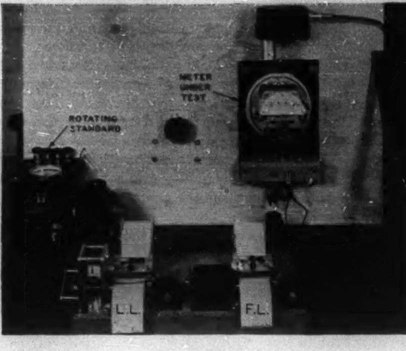
1. Changes in calibration of a meter during a warming-up period.
2. Changes in calibration due to changes in ambient temperature.
3. Variations in register and bearing friction when the meter under test is operating at light load.

It also is possible to evaluate the effective variations in the speed of the standard as determined from the drum driven by the synchronous motor. This is done by using an ordinary meter in the test position and adjusting it to close agreement with the standard, revolution for revolution. If now, the meter so adjusted is slowed down until the pattern again approximates a straight horizontal line, the drum would have relatively increased its speed by one revolution in the interval. This corresponds to an advanced position of the rotating standard disk which progresses uniformly for each pulse, finally making one complete revolution. A study of the variations as indicated by the chart readily permits evaluating the variations in the rotating standard speed. A portion of such a chart is illustrated in Figure 7, whereas Figure 8 illustrates a portion of a chart showing a meter initially 0.25 per cent slow, then correct and finally 0.25 per cent fast. Figure 9 shows a pair of experimental recorders, used for full load and light load testing.

Other arrangements are equally possible and under certain conditions might

Figure 8. Section of tape with overlaid protractor obtained during period of meter adjustment





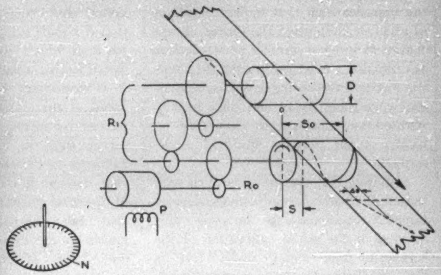


Figure 9 (left). Experimental setup for meter testing and calibrating using separate comparators for full load and light load

Figure 10 (above). Line schematic diagram of most general arrangement of comparator mechanism

be advantageous. For example, it is possible to provide the rotating standard, say with two photoelectric systems so arranged as to constitute, with their amplifiers, a substantially perfect 2-phase system. These voltages could then be applied to windings or plates of a cathode-ray oscillograph tube with a resulting circular pattern. The impulses from the meter under test could then be used to mark the position of the electron beam so as to leave an image which would persist for a short period of time. Its position could then be compared with the next succeeding position and from this the accuracy of the meter under test could be determined. This, while not leaving a permanent record, might have certain obvious advantages.

The utilization of automatic recording devices described above should, in a large degree, reduce personal errors in the testing of watthour meters in the laboratory to an extent which would permit them to be classed as precision measurements. The final results will depend not only upon the accuracy to which a standard has been calibrated but on its reliability and freedom from error under changes in load, frequency, power factor, and temperature which are known to exist.

Category III using indicating instruments in which the measurement of time is not directly involved offers considerable promise in respect to improved accuracy in the calibration of watthour meters. This comprises in incorporating into or as a part of the indicating instrument or the a-c/d-c comparator a means for generating a frequency which is directly proportional to watts.

A number of methods are available for accomplishing such results. As applied to torsion-head type dynamometers, for example, a variable air capacitor could be associated with the torsion spring shaft which, in combination with suitable electronic circuits, will generate (over a given range) a frequency directly proportional to watts.

This generated frequency could then be used for driving the comparator in the manner indicated. In this event we would have:

$$f=K_0W \text{ and } \frac{1}{K_0}\int_{t_0}^{t_1} f_1dt=\int_{t_0}^{t_1} W_1dt$$

The expression on the right is simply the energy in watt-seconds over the interval (t_1-t_0) and on the left, a constant times the number of cycles counted over the same interval. So far as is known no devices of this character have, as yet, been developed.

Appendix

The comparator in its most general form is illustrated in Figure 10 and the following constants and symbols appear in the equations which define its operation:

t = time in seconds
K_s = watt second constant of the rotating standard
K_d = watt second constant of the motor driven drum
K_t = watt second constant of the meter under test
W = load in watts on the standard and the meter under test
N_0 = speed of the standard in revolutions per second
N = number of peripheral slots in the disk of the standard
P = number of pairs of poles on the stator of motor driving the drum
R_0 = ratio of reduction between motor and drum
R = ratio of reduction between drum shaft and paper drive
f = frequency in cycles per second generated in photocell of standard
S_0 = pitch of the spiral on drum expressed in inches
S = distance along drum of the spiral expressed in inches
ΔS = distance between dots as measured in inches across the tape
$\alpha=\Delta S/S_0$ = distance ΔS expressed as a fraction of a drum revolution
N_s = speed of the synchronous motor in rps
N_d = speed of the drum in rps
n_d = revolutions of the drum in some time interval t
n_t = revolutions of the meter under test
n_s = revolutions of the standard
h = distance as measured along the tape
E = error, in per cent, of the meter under test
D = diameter of paper drive drum in inches

We have then

$$f=N_0N=\frac{W}{K_s}N \qquad (2)$$

$$N_s=\frac{f}{P}=\frac{WN}{K_sP} \qquad (3)$$

$$N=\frac{N_s}{R_0}=\frac{WN}{K_sP\,R_0} \qquad (4)$$

When testing a watthour meter by means of another taken as a standard it is customary to express the error in per cent by the following formula:

$$E=100\,\frac{K_tn_t-K_sn_s}{K_sn_s} \qquad (5)$$

This same formula is used in the comparator application, except in the latter the watt-second constant of the drum K_d is substituted for K_s and drum revolutions $n_d+\alpha$ are substituted for n_s. Here, the sign of α is positive if the successive positions of the drum at the instant of recording are in the direction of its rotating and negative if not. With this understanding, the error of the meter under test referred to the standard is,

$$E=100\,\frac{K_tn_t-K_d(n_d+\alpha)}{K_d(n_d+\alpha)} \qquad (6)$$

Where n_d represents integral revolutions of the drum and the fractional part of a revolution while the meter under test makes n_t revolutions. We assume further that the meter under test is in adjustment when

$E=0$ and $\alpha=0$. This makes $K_t n_t = n_d K_d$, from which n_d, the number of drum revolutions required for correct registration may be computed. In the most general case illustrated in Figure 10,

$$n_d=\frac{K_t n_t N}{K_s P R}$$

For any deviation from this, the error becomes

$$E=-\frac{100\alpha}{n_d+\alpha} \qquad (7)$$

If α is small we may write without appreciable error

$$E=-\frac{100\alpha}{n_d} \qquad (8)$$

We also have the following relations:

$$\frac{dS}{dt}=S_0 N_d=S_0\,\frac{WN}{K_s P\,R_0} \qquad (9)$$

Where S generally indicates the intersection of the spiral ridge with a line along the surface of the cylinder parallel to the axis, and measured from some starting position 0, as shown in Figure 10.

If R_1 is the ratio of reduction between the drum and paper drive, and D is the diameter of the drive, we have

$$\frac{dh}{dt}=\frac{\pi D}{R_1}N_d \qquad (10)$$

measured along the paper in inches (travel). This gives, with equation 9

$$\frac{dS}{dh}=\frac{S_0 R_1}{\pi D} \qquad (11)$$

and,

$$h=\frac{S}{S_0}\cdot\frac{\pi D}{R_1} \qquad (12)$$

where S now is the total distance traveled by the intersection parallel to the axis of the cylinder. This may further be expressed in terms of drum revolutions, and we have,

$$\frac{S}{S_0}=n_d+\alpha$$

$$h=(n_d+\alpha)\frac{\pi D}{R_1} \qquad (13)$$

As a practical example, let us make up a combination suitable for testing a meter at full load. First, let us assign the following values to some of the constants

$N=180$; $R_0=1$; $p=3$, and $S_0=2.0$

For one revolution of the master meter we have

$$n_d=\frac{N}{R_0 P}=\frac{180}{3}=60 \text{ revolutions}$$

If we further assume $K_t=K_s$ then $E=0$ when $n_d=60$ and $\alpha=0$. Since $h=(n_d+\alpha)\pi D/R_1$, the choice of D and R_1 will determine the value of h for any given value $(n_d+\alpha)$. Let us choose them in such a way that the distance traveled by the chart for $(n_d+\alpha)$ revolutions of the drum is equal to the distance ΔS between dots corresponding to successive disk revolutions for $E=-0.1$ per cent.

From equation 7 we find that $n_d+\alpha=1{,}000\alpha$, or

$$h=\frac{1{,}000\alpha\pi D}{R_1}$$

now

$$\alpha=\frac{\Delta S}{S_0} \text{ so}$$

$$h=\frac{500\Delta S\pi D}{R_1}$$

a suitable diameter for D would be about 1 inch and since we have decided to make h equal to ΔS

$R=500\pi$ or approximately 1,571

In most instances the approximate formula 8 is sufficiently accurate for determining E from

$$E=-\frac{100\alpha}{n_d}=-0.833\Delta S$$

with the same meaning for ΔS. If, under steady conditions the dots arrange themselves in a straight line inclined at an angle θ to the direction of the tape, we have,

$$\tan\theta=\frac{\Delta S}{h}$$

This may be expressed in terms of E using equations 13 and 7 and observing that $\alpha=\Delta S/S_0$. From this we get,

$$\tan\theta=\frac{\Delta S}{h}=-\frac{ES_0R_1}{100D\pi}$$

on substituting the above values, $\tan\theta=-10E$. For the light load testing of watthour meters, it is desirable to operate the master meter at full load speed. This is particularly true in the comparator since it requires practically no change in the electronic components and as in other forms of testing equipment the master meter need only be calibrated at one load.

As previously pointed out, however, the above comparator operating under such condition would cover a range of errors too small to be of any practical value. To illustrate, n_d would then become 600 since the master meter would make ten revolutions to one of the meter under test. The error in terms of ΔS would be $E=-0.0833\,\Delta S$ and 0.12 inch between dots would be only 0.01 per cent.

To avoid this difficulty it is necessary to make use of a second recorder having different constants. In the present design the following constants were used

$N=180$; $R=300$; $D=1.0$; $P=8$; $S_0=2.0$;
$R_0=11.25$

From this we have,

$E=-2.5\Delta S$
$\tan\theta=-1.91E$

References

1. Establishment and Maintenance of the Electrical Units, **F. B. Silsbee.** *Circular 475*, National Bureau of Standards (Washington, D. C.), June 30, 1949.

2. Standard Electro-dynamic Wattmeter and AC-DC Transfer Instrument, **J. H. Parks, Arthur B. Lewis.** *Research Paper RP-1244*, National Bureau of Standards (Washington, D. C.), volume 25, November 1940.

3. A Precision AC-DC Comparator for Power and Voltage Measurements, **G. F. Shotter, H. D. Hawkes.** *Journal*, Institution of Electrical Engineers (London, Engand), volume 93, number 34, August 1946.

4. The Accuracy of Commercial Electrical Measurements, **H. B. Brooks.** *AIEE Transactions*, volume 39, part I, 1920, pages 495–615.

5. A Precision Rotating Standard for the Measurement of Kilowatt-Hours, **J. H. Goss, A. Hansen, Jr.** *Electrical Engineering* (*AIEE Transactions*), volume 59, July 1940, pages 412–16.

6. A New Precision A-C Voltage Stabilizer, **G. H. Patchett.** *Journal*, Institution of Electrical Engineers (London, England), volume 97, 1950, part II, page 529.

7. Photo Electric Equipment for Testing Watthour Meters. *Bulletin G.E.H. 1240-A*, General Electric Company (Schenectady, N. Y.).

8. Improvement in Modern Meter Testing Technique, **E. A. Lynch, M. A. Princi.** *Electrical Engineering* (*AIEE Transactions*), volume 61, April 1942, pages 218–223.

9. Recent Developments in Comparative Methods of Testing A-C Electricity Meters, **S. F. Musson, R. E. Mell.** *Journal*, Institution of Electrical Engineers (London, England), volume 97, Part II, April 1950.

Discussion

F. B. Silsbee (Bureau of Standards, Washington, D.C.): A large part of the value of this excellent paper arises from the broad point of view with which the author has approached the problem. His classification of test methods into three categories is very helpful.

The development of Type II methods results primarily from the fact that the watthour meter has possibilities as a standard for maintaining the unit of energy, which can challenge the best performance of the wattmeter used to calibrate it. This constancy makes it possible as a standard for measurements in the range of accuracy approaching 0.01 per cent. This was shown by the pioneer work of Agnew[1] and more particularly of Goss and Hansen, see reference 5 of the paper. The development of Type II methods based upon such energy standards has taken place mainly in two rather different directions:

1. Stroboscopic procedures for field or laboratory tests of large numbers of meters rapidly and with moderate accuracy.
2. Procedures such as Holtz's new watthour meter comparator for very precise laboratory calibrations and studies.

Work on an alternative precise comparator in the second item under category II was started at the National Bureau of Standards in 1940 but a War and an Emergency have diverted so much man power to military problems that we are even yet not in a position to state the accuracy attainable, although preliminary trials look promising. The National Bureau of Standards device is closely analogous to Holtz's in its purpose, but quite different in its method. In it a circular phosphorescent scale is traversed by a spot of light reflected from a mirror on the shaft of the standard meter. At the beginning and end of any desired number of revolutions of the meter under comparison, the light is momentarily intensified leaving a transient bright spot. The shift in successive locations of this spot relative to the fixed scale indicates the difference in rate of the two meters.

Holtz's category III may in the long run turn out to hold the greatest value for the future. The idea of counting electrical cycles rather than the revolutions or oscillations of a ponderable body has long seemed attractive because it completely eliminates the friction which unavoidably affects mechanical integrators. For a long time it was thought that the disturbing effects of the mechanical reaction of a frequency-generating circuit on the torque developed by the power-measuring circuit would present an unsurmountable obstacle. Nowadays, with the magic of electronics, it seems possible that the use of a servosystem for automatically adjusting a frequency to be proportional to the power without reacting appreciably on the wattmeter movement, or of a modern high-speed counting circuit may offer a chance for a revolutionary change in metering practice.

Assuming freedom from relay or other switching errors, it would appear that the variation of the disk speed with angular position might well set a limit to the accuracy of rate determinations in short-time runs. It is therefore important to know the magnitude of variations of disk speed with position for conditions of constant power. We would like to inquire whether Mr. Holtz has any data available concerning this variation and also whether his method might permit a separation of the effect resulting from the gear-train and from eccentricity or nonuniformity of the disk itself.

REFERENCE

1. A WATTHOUR METER METHOD OF TESTING INSTRUMENT TRANSFORMERS, **Agnew.** *S233*, National Bureau of Standards (Washington, D. C.), Volume II, 1915, page 347.

J. M. Vanderleck (The Hydro-Electric Power Commission of Ontario, Toronto, Canada): The vast improvements which have been made in the single-phase watthour meter since the appearance of the early types at the beginning of the century, have been brought about largely through the co-operative efforts of makers and users. No small part of the effort has been the continuous testing and investigation of watthour meters in order to devise meters with improved performance. Clearly, further progress in the performance of both standard and service-type watthour meters depends a great deal on continuation of investigation of the causes and elimination of meter errors. Today, some lines of investigation are limited by the precision with which tests can be carried out conveniently. The great merit of the author's novel watthour meter comparator is that it enables precise measurements to be made in the laboratory much more conveniently than has been possible heretofore.

The author has shown that by further development, the recording equipment may have its usefulness extended to the direct calibration of watthour meters using suitable ac-dc transfer instruments. This requires producing a frequency proportional to watts. If the relationship between frequency and watts is to be constant within 0.01 per cent, for example, it is questionable whether it may not be simpler to improve existing apparatus to hold watts to this degree of constancy. Then a constant, crystal-controlled frequency can be supplied readily to the synchronous motor driving the drum, and the desired accuracy of measurement is achieved.

It is inherently advantageous to hold watts constant, for in this case the electromechanical transient performance of the watthour meter comparator does not enter into the over-all accuracy of test. I should like to ask the author if it is satisfactory to use the commercial a-c service to supply the standard and test meters, and if not, how smooth the supply must be to avoid errors from transient effects, such as the lag in phase position or slipping of the synchronous motor after a sudden change in meter disk speed. If the synchronous motor were a type which could hunt or shift phase by ±180 electrical degrees without slipping a pole, then discrepancies of ±0.3 per cent could be experienced when the disk markings are 180 in number. It would be interesting to know if this type of error has been eliminated in the apparatus described.

The author states that in the well-known direct method of test, where load is held constant, it is more accurate to measure the elapsed time for an integral number of revolutions of the meter under test than to start and stop the meter and measure the revolutions for an elapsed integral number of seconds. The truth of this statement depends upon the point of view. In carrying out accuracy checks on portable standard watthour meters, it is considered to be more accurate to start and stop the standard meter being checked because in normal use the accepted procedure is to start and stop the portable standard meter.

As for the cause of start and stop errors, even with no effective friction, it does not seem that the loss in rotation of the moving system during the period of starting will be exactly compensated for by a gain in rotation after the current circuit of the meter has been broken, because the damping effect of the magnetic flux from the current element is present during the starting period and absent during the stopping period. The effect may be insignificant, but a portable standard watthour meter is normally controlled by the potential circuit, in which case, the discrepancy should be more noticeable because of the higher magnetic flux and damping effect from the potential element.

The author has classified the new watthour meter comparator under the third group listed in category II in which the relative rotations of the standard and test meters are automatically recorded on a strip chart. In the same group could be mentioned the well-known chronograph method of comparison, where each revolution or half revolution of both meters is recorded on a strip chart.

When an automatic permanent record of the comparison of a standard and test meter is not required, but a high order of precision is called for, a proposed simple method is to use a crystal-controlled oscillator and two electronic decade counters as timers, one for each meter. Using the anticreep holes, photoelectric impulses from the two meters can be arranged to switch the two timers electronically for any integral number of half revolutions. An accuracy of 0.01 per cent should be realized by holding load to 0.1 per cent, arranging the test runs on both meters to be coincident within 10 per cent, and operating the decade counters from a timing frequency of 10,000 cycles. This scheme should be capable of comparable accuracy to the author's watthour meter comparator, and avoids the necessity of cutting slots in the disk of the standard meter.

B. E. Lenehan (Westinghouse Electric Corporation, Newark, N. J.): This paper covers the field of meter testing methods quite completely in general terms. I agree with the advantages and disadvantages given. This device should be quite useful for studying abnormal performance of occasional meters and for checking performance of standard meters. The performance of the standard is the limiting factor in the overall accuracy.

One additional point may be added to the observations about start-stop test methods. Power in a single-phase circuit pulsates at 120 cycles with a variation in energy content of two-thirds the average rate. If the time of test involves an odd quarter-cycle of power-line frequency, a variation in reading obtained will result. If one desired to make a full-load test on a 30-rpm meter for one revolution, 480 quarter-cycles would be the time, and a variation of ± two-thirds of 1/480 could result, or a total range of one part in 360. Consequently, any rapid test method should be of the continuously running type.

F. H. Busch (General Electric Company, Lynn, Mass.): The problem of testing a-c watthour meters is indeed a special one, challenging to both meter manufacturers and users. The object of testing is to achieve a degree of accuracy in the calibration of watthour meters that is consistent with their application, satisfactory to regulatory bodies, and takes full advantage of the reliability of the meters themselves. The ultimate objective is, of course, the lowest possible expense in metering electrical energy. Careful examination of the methods and apparatus for meter testing, such as Mr. Holtz has made, is important if progress is to be made toward this objective.

The special type of chronograph which Mr. Holtz describes is interesting, and appears capable of testing precision suitable for the most critical testing problem. The ease of reading of the record chart is a commendable feature, although the limited range of accuracy difference which the comparator can measure does limit or complicate its application to many laboratory problems. Of course, changing the gear ratios can extend the range at a compromise in the precision of the chart.

It would appear that the comparator

might find application to the calibration of large quantities of meters, which would help to justify the original and maintenance expense of the rather complex equipment. The calibration of new watthour meters by a meter manufacturer is such an application. It would be of interest to know if the author's work with this testing method has enabled him to appraise its possibilities for such factory calibrations.

G. B. Schleicher (Philadelphia Electric Company, Philadelphia, Pa.): The watthour meter comparator described by Mr. Holtz is a step forward in improving the accuracy with which a watthour meter and a rotating standard may be compared. It appears that an instrument such as this would be useful primarily in the laboratory of the meter manufacturer, where load and voltage conditions may be controlled accurately and the load may be supplied from a sine-wave generator. Devices of this type may be adapted by the manufacturer for the study of improvements in his product, such as eliminating the sources of variations in speed for various positions of the disk, which would be useful for stroboscopic testing; and also to study friction effects such as may occur when different parts of the register gearing are in mesh.

While the modern watthour meter has been developed to have very good performance, the range of deviations between a test under controlled conditions in the laboratory and its performance in the field under variations in voltage, load, power factor, temperature, and harmonics which may vary at different parts of the system and even at different times of the day, may be of the order of as much as ±0.5 per cent. To this must be added the possible deviation of the rotating standard recognized in the paper as ±0.2 per cent. Since the rotating standard is fundamentally a watthour meter, its errors will tend to be similar to those of the meter under test.

For routine use in the testing of meters by public utilities, where meters are necessarily tested on the system, a device capable of reading errors of the order of 0.01 per cent, appears to give a misleading sense of accuracy. As pointed out by the author, the speed of the drum may be modified, but for routine use by meter testers, a definite indication of the results in steps of 0.1 per cent would be more useful than the use of an overlaid protractor. Undoubtedly, Mr. Holtz's development will lead to further improvements in meter testing methods.

G. G. Waite (Sangano Company Limited, Leaside, Ontario, Canada): The presentation of a paper on an integrating meter subject is a rare and welcome event. When the author is as distinguished in his field as is Mr. Holtz, it is doubly welcome.

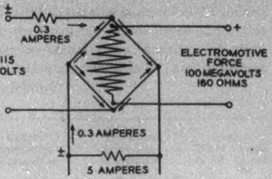


Figure 1

I would like to supplement some of Mr. Holtz's remarks on the subject of transfer standards. Ordinary dynamometer wattmeters, such as the Weston Model *310*, are inadequate, as any integrating meter man will agree. The long scale instruments have usually the same movement, and while the reading error is reduced, portability is lost.

What is needed is a really rugged, portable, easily checked standard of watts that preferably has no moving parts, is by basic principle free from frequency effects up through, say, audio frequencies, can be shipped without particular care anywhere, has errors of the same order as the Shotter-Hawkes device and a scale length that practically eliminates reading uncertainty.

This instrument is not imaginary, but is potentially available today. An indirectly heated thermal convertor can be readily modified to operate on alternating or direct current by a suitable resistance network. The standard indirectly heated thermal convertor now being widely used in Canada has the following constants:

1. For the 2-element unit structure the output is 100 megavolts maximum for 5 amperes.
2. One hundred volts per element input.
3. Ninety per cent of ultimate output in three-fourths second.
4. Current burden 2.9 watts at 5 amperes.
5. Voltage burden $10^{1}/_{2}$ volt-amperes, 0.66 power factor at 115 volts, about 400 junctions and 36 ohms resistance in the couple circuit.

The modified instrument, which I will call the ac-dc thermal wattmeter, has about the same electromotive force output and time of response, but is a single element rather than a polyphase element. Due to the necessity of impedance matching by pure resistance the burdens are enhanced, and the current circuit and voltage circuit must be supplied from separately insulated sources.

Live load can be read on alternating current if a potential or current transformer is used to insulate between the voltage and current circuits. Two separate batteries are required for direct current.

At 115 volts the voltage burden is 35 watts and at 5 amperes the current burden is 30 watts. The output is, of course, read on a potentiometer so reading error can be negligible. The ac-dc thermal wattmeter has over 2,600 junctions and 160 ohms resistance in the couple circuit.

F. C. Holtz: The author wishes to thank those who have contributed so generously to the discussion of this paper. It is hoped that the interest shown in the subject will encourage the publication of new material.

In answer to Dr. Silsbee I should like to say that the detection and measurement of variations in the speed of the disk of the rotating standard with respect to position for the condition of constant power is possible. One method is outlined in the paper which even permits some variation in the applied power. It is not possible, however, to separate these variations into components representing variations in mechanical friction and variations from all other factors. The suggested method is furthermore confined to relatively high speeds where variations in mechanical friction are only a very small percentage of the total. Unless a method is devised for determining these quantities at low rotational speeds, its usefulness is limited.

In reference to Mr. Vanderlick's remarks the answer will depend upon the degree of accuracy desired in the final results. For precision laboratory measurements a comparator of relatively high resolving power should be used. In such instances a well regulated a-c power supply must be used. This is especially necessary when calibrating the rotating standard by means of the ac/dc comparator or other highly sensitive a-c wattmeter. For the commercial testing and calibrating of single phase a-c watthour meters experience indicates that commercial a-c service is generally sufficient. No study has been made of the errors introduced through sudden power variations such as might cause hunting of the synchronous motor. The proper choice of synchronous motor will, in a large measure, minimize the errors from such causes.

In reference to the remarks of Mr. Busch the author wishes to state that a watthour meter comparator similar to that described in the paper has been used for checking the calibration of watthour meters from regular production. Experience indicates that it has distinct possibilities as a device for the regular factory calibration of watthour meters. In such instances, however, the resolving power would be somewhat decreased in accordance with commercial requirements.

Mr. Waite has given considerable encouragement to the possibility of using thermal converters as transfer instruments in the measurement of a-c power. If such devices could be manufactured relatively free from error they would have the distinct advantage of having no moving parts and therefore less subject to derangement from handling.

Tubes for Dielectric Heating at 915 Megacycles

RICHARD B. NELSON
NONMEMBER AIEE

Synopsis: Dielectric heating can be much faster at microwave frequencies. Because leakage radiation is large enough to interfere with communications, heating is best done at assigned frequencies. The bands at 915 and 2,450 megacycles are most useful. An industrial magnetron oscillator generates 5 kw at 915 megacycles. The tube and associated control circuits are described. Apparatus for heating dinners and plastics has been constructed using this tube. Experimental magnetrons give 30 kw at 1,000 megacycles.

MICROWAVE tube developments several years ago reached the stage where kilowatts of power were available at frequencies of 10^9 to 10^{10} cycles. The question quickly arose "Why cannot all dielectric heating which is difficult or impossible at low frequency be done easily at these microwave frequencies?" Fundamental to dielectric heating is the relationship

$$P = 2\pi f E^2 \epsilon'' \times 0.0885 \times 10^{-12} \qquad (1)$$

where

P = power in watts per cubic centimeter converted to heat in the dielectric
f = frequency of alternating electric field
E = rms field strength in volts per centimeter
ϵ'' = loss factor of the dielectric

The rate of heating for any material is thus:

$$\frac{dT}{dt} = \frac{P}{4.186C} = \frac{1}{C} 2\pi f E^2 \epsilon'' \times 0.0211 \times 10^{-12} \qquad (2)$$

where

dT/dt = temperature rise in degrees centigrade per second
C = specific heat of the material

If it is desired to increase the rate of heating a given dielectric material, the field strength or the frequency may be increased. The field strength is easy to control, but it can be increased only to the limit imposed by arcing between the electrodes and the work. The filed strength at which arcing occurs varies tremendously with different kinds of materials. For a box of raisins—high-dielectric-constant lumps touching at small points—there is a concentration of field at the contact points which causes arcing and local burning at low electrode voltage.

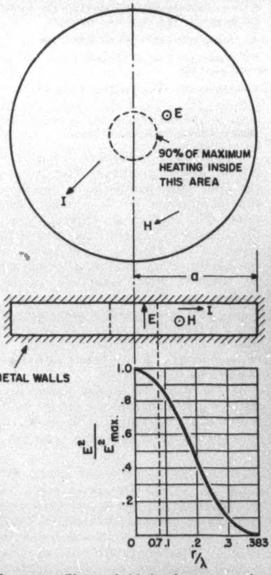


Figure 1. Electric field distribution in cylindrical resonator. Rate of dielectric heating at any point is proportional to E²

Choice of Frequency

It is seen from equation 2 that the rate of heating at constant voltage is proportional to frequency if the loss factor does not change. In most materials, the loss factor actually rises with frequency in the useful ranges. This sounds like an unmitigated blessing: a sufficiently high frequency can be used to heat any material as fast as is desired with a voltage safely below the breakdown point. There are, however, limits on usable frequencies imposed by economic and, particularly, physical factors.

FREQUENCY ALLOCATIONS

In the microwave region, it has been found extremely difficult to shield dielectric heating equipment to reduce radiation below the Federal Communications Commission's allowable maximum. For example, a series of tests were made of

Figure 2. A 5-kw magnetron tube. The tube on the right is completely sectioned, except the cathode

---

Paper 52-66, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted December 1, 1950; made available for printing December 12, 1951.

RICHARD B. NELSON was with the General Electric Research Laboratory, Schenectady, N. Y., when this paper was written. He is now with Varian Associates, San Carlos, Calif.

radiation from a resonant cavity box. The cavity contained a dinner and was fed with 4 kw of power at 1,040 megacycles. The following average radiated field strengths were measured.

Method of Closing Cavity	Radiated Field Strength, Microvolts Per Meter at 1 Mile
Radio-frequency choke filter	1,000
Finger contacts	300
Choke filter plus surrounding metal cabinet	400
Soldered cover	small

These measurements, made by Dr. H. W. Anderson of the General Electric Electronics Laboratory, are not presented as accurate data. Their order of magnitude demonstrates, however, that shielding is impractical in the microwave region.

Heating apparatus is essentially confined to the frequency bands assigned for this service where large amounts of stray radiation are permitted. These bands are as follows:

Center Frequency, Megacycles	Deviation Allowance, Megacycles	Quarter Wave Length in Air, Inches
915	± 25	3.22
2,450	± 50	1.21
5,850	± 75	0.50
10,600	±100	0.28
18,000	±150	0.16

AVAILABILITY OF POWER SOURCES

Standard grid-controlled vacuum tubes serve as efficient oscillators up to the order of 100 megacycles per second. Their wide use in communications has made them highly developed and cheap.

Throughout the range between 100 and 900 megacycles, vacuum-tube oscillators can be built to give as much power as probably is needed. These are specialized tubes in small production and expensive. In view of the lack of frequency assignments, it is doubtful if these frequencies will be used much for heating.

At 915 megacycles, a rugged magnetron oscillator, to be described in this paper, is available to give 5-kw output. Experimental magnetrons have been built to produce 50 kw in this frequency range.

At 2,450 megacycles, magnetrons are available from the Raytheon Manufacturing Company giving up to 2-kw output.

At higher frequencies, tubes suitable for industrial use have not yet been put on the market. They can be made if a demand for them develops.

UNIFORMITY OF HEATING

Most of the materials to be heated with microwaves are relatively low-loss dielectrics. This means that in the passage of an electromagnetic wave through the material, only a small part of the energy in the wave is absorbed per wave length of travel. If it is desired to dissipate all the energy in the wave in a piece of dielectric no more than a few wave lengths long, the wave must be reflected back and forth

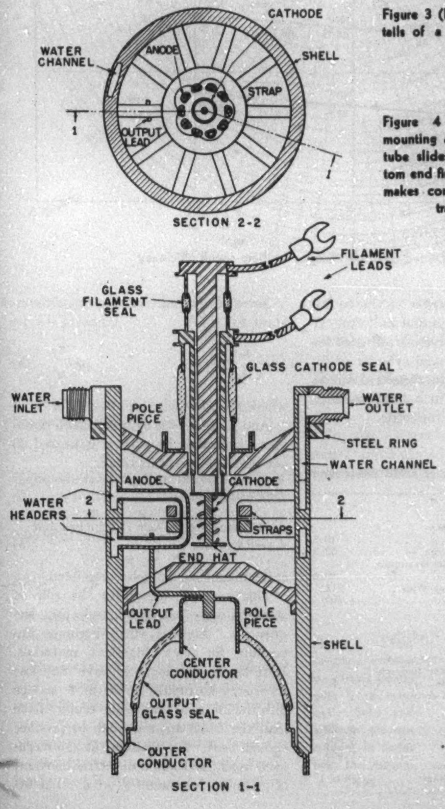
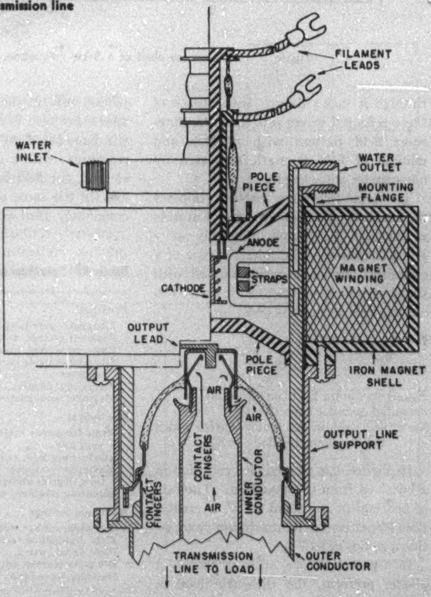

CATHODE
WATER CHANNEL
ANODE
SHELL
STRAP
1
OUTPUT LEAD
1
SECTION 2-2
FILAMENT LEADS
GLASS FILAMENT SEAL
GLASS CATHODE SEAL
WATER INLET
POLE PIECE
WATER OUTLET
STEEL RING
WATER CHANNEL
ANODE
CATHODE
WATER HEADERS
2
2
STRAPS
END HAT
OUTPUT LEAD
POLE PIECE
SHELL
CENTER CONDUCTOR
OUTPUT GLASS SEAL
OUTER CONDUCTOR
SECTION 1-1

Figure 3 (left). Construction details of a 5-kw 915-megacycle magnetron

Figure 4 (right). Magnetron mounting and connections. The tube slides into the magnet, bottom end first, and simultaneously makes contact with the output transmission line


FILAMENT LEADS
WATER INLET
WATER OUTLET
POLE PIECE
MOUNTING FLANGE
ANODE
MAGNET WINDING
STRAPS
CATHODE
OUTPUT LEAD
POLE PIECE
IRON MAGNET SHELL
AIR
AIR
OUTPUT LINE SUPPORT
CONTACT FINGERS
CONTACT FINGERS
INNER CONDUCTOR
AIR
TRANSMISSION LINE TO LOAD
OUTER CONDUCTOR

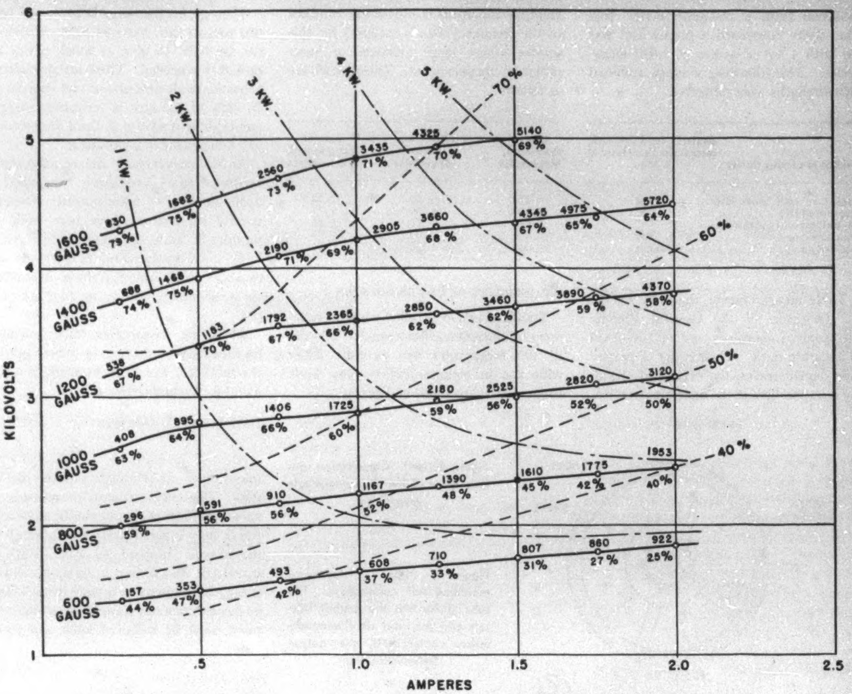


Figure 5. Performance chart of a 5-kw magnetron. Dotted lines are contours of plate-circuit efficiency

through it many times. Interference of these reflected waves sets up a standing-wave field pattern with maxima and minima of field strength separated by one-quarter wave length.

The standing-wave pattern imposes definite restrictions on the size of objects which may be heated uniformly at a given frequency. Consider the case of a cylindrical resonant cavity excited with the electric field parallel to the axis. The electric field strength is given by

$$E = E_{\max} J_0(6.28r/\lambda) \tag{3}$$

where

$E_{\max}$ = maximum field (on the axis)
r = radial distance from the axis
λ = wave length in the material with which the resonator is filled

In Figure 1 is plotted the variation in E^2 derived from this equation. The rate of heating, proportional to E^2, is greater than 90 per cent of its maximum value inside a radius of 0.07 λ.

In a practical case, such as heating a plastic preform, the dielectric load is placed only in the center of the cavity where the field is high and uniform. If the diameter of the cavity is adjusted for resonance, with a coaxial cylinder of dielectric, the field in the dielectric will be exactly the same as in a smaller cavity completely filled with dielectric.

Table I. Specifications of *RN48* Magnetron

Electrical	
Filament—pure tungsten	
Filament voltage, volts	10.5
Filament current, maximum, amperes	53.5
Filament starting current, maximum, amperes	80.0
Frequency, nominal, megacycles	915
Magnetic field, gausses	1,400
Mechanical	
Radio-frequency output, $3^1/_8$-inch coaxial line	
Anode water flow, gallons per minute	2
Maximum outgoing water temperature, degrees centigrade	70
Mounting position—axis vertical	
Maximum Ratings	
D-c plate voltage, volts	5,000
Plate dissipation, watts	3,000
Plate input, watts	7,500
D-c plate current, amperes	2
Standing wave ratio, voltage	3 to 1

For a plastic material of dielectric constant 4.0,

$$\lambda = \frac{\lambda_0}{\sqrt{4.0}} = 0.5\lambda_0 \tag{4}$$

where λ_0 = wave length in free space.

And if the plastic cylinder has a diameter d such that its surface is heated 90 per cent as much as its center

$$d = 2 \times 0.07 \times 0.5\lambda_0$$

At 915 megacycle, $\lambda_0 = 12.9$ inches and
$d_{915} = 0.9$ inches (5)

is the maximum diameter permitted.

This example illustrates the role of standing-wave patterns in selecting frequencies. Another effect becomes important for very high-loss materials. Here the electromagnetic wave may lose its energy so rapidly in its first passage into the material that at the center of the load the fields are too weak to produce enough heat. For example, at 915 megacycle roast beef has a dielectric constant of 28 and a loss factor $\epsilon'' = 5.6$. The at-

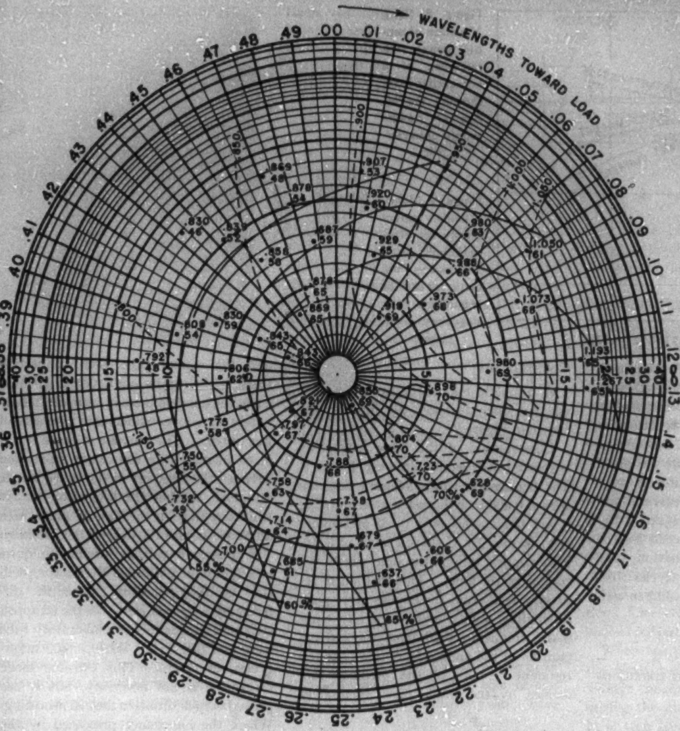


Figure 6. Rieke diagram showing variation of wave length and efficiency with load impedance. The radial co-ordinate is voltage reflection coefficient of the load on the transmission line. (Scale marked in standing wave ratio in decibels.) Azimuthal co-ordinate is phase angle of the reflection coefficient. (Scale marked in wave lengths displacement of standing wave minimum.) Dotted lines are constant wave length contours, numbered in centimeters above a fixed value. Solid lines are constant efficiency contours. Data taken at 1.0 ampere plate current, 1,200 gausses

tenuation constant α for the fields of a wave is given by

$$\alpha = \frac{\pi}{\lambda_0}\frac{\epsilon''}{\sqrt{\epsilon'}} \tag{6}$$

where ϵ' = real part of dielectric constant. The power density in the wave,

$$E^2 = E_0^2 e^{-2\alpha} \tag{7}$$

If the effective depth of penetration p of the wave is defined as that depth at which the power density has fallen to $1/\sqrt{e}$ or 0.61 of its initial surface value.

$$p = \frac{1}{4\alpha} = \frac{\lambda_0\sqrt{\epsilon'}}{4\pi\epsilon''} = \frac{\lambda}{4\pi\epsilon''} \tag{8}$$

In roast beef, $p = 0.97$ inches effective penetration at 915 megacycles.

These examples indicate how, for any type of load, consideration of the wave length in the material in relation to the size of the piece leads to the choice of frequency. Present applications use the 2,450-megacycle and the 915-megacycle bands, and it appears that these will remain the most important in the microwave region, even when tubes become available for higher frequencies.

The 915-Megacycle 5-Kw Magnetron

In the General Electric Company, microwave heating investigations and developments have been concentrated at

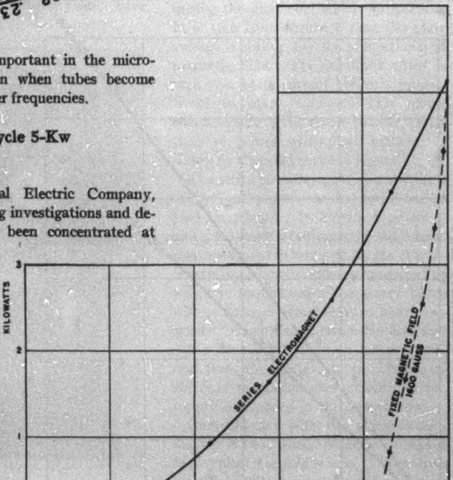


Figure 7. Effect of plate voltage on output power with fixed magnetic field and with an electromagnet in series with the magnetron

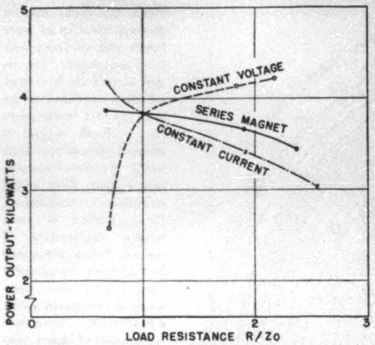
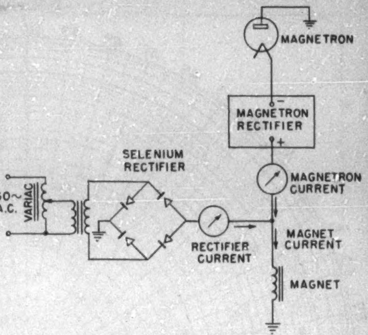


Figure 8 (left). Variation of output power with resistance of a nonreactive load, showing the effect of power-supply characteristics



MAGNETRON
MAGNETRON RECTIFIER
−
+
SELENIUM RECTIFIER
MAGNETRON CURRENT
60~ A.C.
VARIAC
RECTIFIER CURRENT
MAGNET CURRENT
MAGNET

Figure 9 (right). Auxiliary current supply for electromagnet, used for starting the oscillator and for controlling the magnetron plate current

915 megacycles. Tube development, which had to come first, was started in the General Electric Research Laboratory in 1945 with a 5-kw magnetron oscillator, the *RN48*, operating at 1,040 megacycles. The factory model of this tube was given the developmental designation of type *Z1391*. Later, when the Federal Communications Commission assigned a frequency band, it was superseded by the developmental type of designation, type *Z1492*, operating at 915 megacycles. All of these tubes are essentially alike in construction and characteristics.

SPECIFICATIONS

Specifications and operating conditions are listed in Table I.

CONSTRUCTION

A complete magnetron, and a similar one cut open to show the interior, are shown in Figure 2.

Construction of the 5-kw magnetron is shown in the section drawing, Figure 3. The filament is a helix of pure tungsten wire of 0.400-inch diameter. Surrounding it are ten anodes on a 0.687-inch diameter, formed by U-shaped loops of 3/16-inch copper tubing through which the cooling water flows. This method of removing heat from the anodes gives good dissipating ability and practically eliminates frequency drift from thermal expansion of the anode structure. The ten resonant circuits of the oscillator are formed by the anode pipes themselves and the copper tube shell connecting them at their outer ends. At their inner ends, the anode pipes are connected alternately by copper straps into two sets of five anodes. The purpose of these straps is to increase the frequency separation between the resonant modes of the anode structure, so that the oscillation always will take place in the desired mode with alternate anodes 180 degrees out of phase. The resonant circuits would be quite loosely coupled without the straps, with the resulting resonant frequencies differing by a few per cent. In the strapped structure, resonances of a laboratory tube were at 1,063, 2,110, 3,040, 3,750 megacycles, et cetera. This extreme mode separation is not necessary, but it has proved very desirable in dielectric heating. Where the impedance presented by the load to the oscillator may change greatly

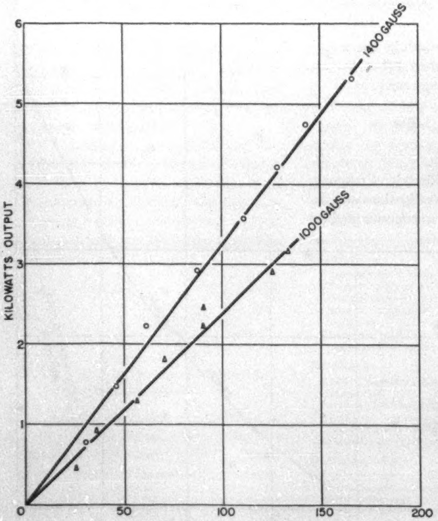
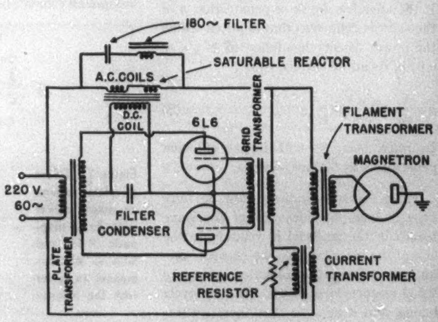


Figure 10 (left). Back heating of the cathode in 5-kw magnetron. These data are for a matched load

Figure 11 (below). Constant temperature filament regulator circuit. Resistance of the magnetron filament is compared with a preset reference. Unbalanced signal is amplified and used to control filament heating current



180~ FILTER
SATURABLE REACTOR
A.C. COILS
D.C. COIL
6L6
GRID TRANSFORMER
FILAMENT TRANSFORMER
MAGNETRON
220 V. 60~
FILTER CONDENSER
PLATE TRANSFORMER
REFERENCE RESISTOR
CURRENT TRANSFORMER

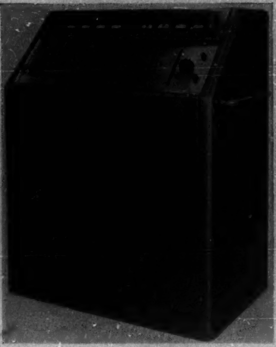


Figure 12. A 915-megacycle heater unit. This oscillator may be used with any loading equipment

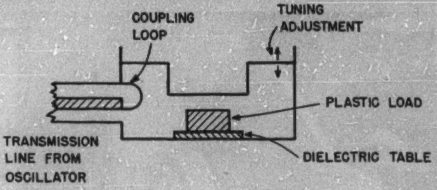


Figure 14 (right). Resonant cavity for heating plastic preforms

it is well to have insurance against oscillations jumping to another mode.

Returning to Figure 3, coupling of the anode tank circuits to the load is provided by a copper strap connected to one leg of one anode pipe and leading through the glass output seal to an external coaxial line. The point of attachment to the anode pipe is determined to give the proper loading of the oscillator when the transmission line is reflectionless.

The tube is mounted by sliding it inside its solenoidal electromagnet. As shown in Figure 4, the magnet has an iron shell which carries the flux. Jumping a short gap formed by the copper tube shell, the flux enters the iron pole pieces inside the tube and provides the magnetic field parallel to the axis of the anode and cathode. As the tube is plugged into its magnet, its output end makes contact with the inner and outer conductors of the transmission line. Air is blown in through the inner conductor to cool the glass seal, which itself undergoes dielectric heating.

PERFORMANCE OF 5-KW MAGNETRON

Operating characteristics of a magnetron are best shown by two sets of curves. Figure 5 is a Performance Chart where contours of magnetic field, output power, and efficiency are plotted on co-ordinates of plate voltage and current. This data is taken with the transmission line matched to the load.

Figure 6 is a Rieke Diagram where, for standard conditions of plate current and magnetic field, the efficiency and wave length are plotted on co-ordinates of reflection co-efficient of the load for wave going down the transmission line.

ASSOCIATED CIRCUITS

Electromagnet Control. The magnetron is essentially a diode, operated for convenience with its anode grounded. The plate power supply is between cathode and ground. We have found it advantageous to return the low-voltage end of this supply to ground through the electromagnet, so the magnet is effectively in series with the plate current of the magnetron. The reason for this may be seen by examination of Figure 5. If the magnetic field is constant, a small change in plate voltage produces a large change in current and output power. The dotted curve in Figure 7 shows this relation. Line-voltage fluctuations can cause serious output variations.

When the plate current flows through the electromagnet, an increase in plate voltage causes increased current to flow, raising the magnetic field in proportion. It is seen from Figure 5 that the plate voltage is almost directly proportional to magnetic field. The resultant effect is such that plate current becomes proportional to plate voltage. This ohmic characteristic gives much smaller variations of power with plate voltage, as shown by the solid curve in Figure 7.

The series electromagnet also stabilizes the output power against variations in load impedance. In Figure 8 power is plotted versus resistance of an ohmic load, measured in terms of Z_0, the characteristic impedance of the transmission line which it terminates. The power curve with a series electromagnet is more constant than with fixed magnetic field and either constant voltage or constant current from the plate supply. Judicious choice of regulation characteristic of the supply could help in this respect also.

One difficulty with the series electromagnet is that when the plate voltage is first applied, the field is zero. Under these conditions, the magnetron will not start in its proper mode but will oscillate in a high-frequency electronic mode at a voltage too low for the proper mode to take

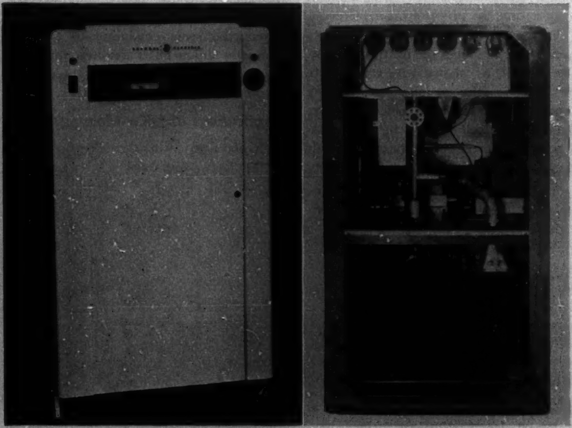


Figure 13. Developmental heater for frozen dinners. Food is placed in the cavity in the front of the machine and a door closes the opening. A frozen dinner is ready to eat in 70 seconds

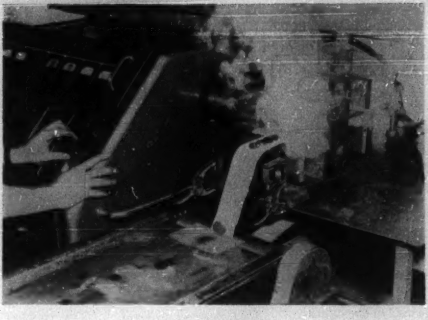


Figure 15. Continuous heating of plastic preforms on a conveyor belt

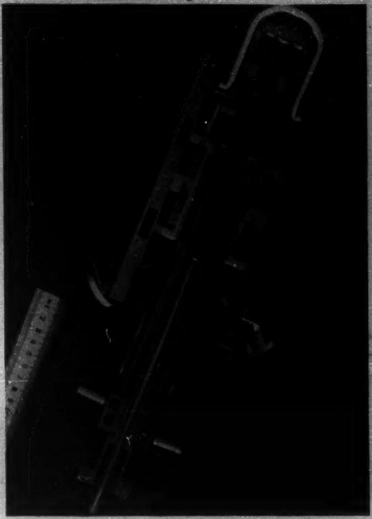


Figure 16 (right). Cut-open view of high-power 1,000-megacycle magnetron. The cathode in this model is a radiation-heated thoria-coated cylinder surrounded by water-cooled copper pipes

over. The remedy for this trouble is to have a residual magnetic field present before the plate voltage is applied. Figure 9 shows a rectifier circuit which supplies starting field by feeding current into the magnet in parallel with the plate current. As the plate current rises, the drop in the magnet exceeds the rectifier voltage and the rectifier stops feeding current.

The rectifier shown in Figure 9 has other uses. It provides a surge path for the magnet current when the plate supply is cut off, avoiding excessive voltage rise due to the inductance of the magnet.

Also, with the variable a-c supply to the rectifier, it may be used as an output power control which is much cheaper than a plate power supply with continuously variable voltage. Leaving plate voltage fixed, the d-c output voltage of the rectifier is raised until it feeds shunt current into the magnet. This raises the magnetic field and reduces magnetron plate current. In this way, complete and continuous power control is obtained with a small Variac.

Filament Temperature Control. One very important piece of circuitry is a device for keeping the magnetron's filament at the proper temperature. The filament receives a considerable amount of energy from bombardment by electrons which get into the radio-frequency field in the wrong phase and are accelerated, just as in a cyclotron. Most of these electrons strike the cathode. In Figure 10 is plotted the back-heating power in a *Z1492*-type magnetron at 1,000 and 1,400 gausses as a function of output power. These data were taken with a load matched to the output transmission line. Unfortunately, the heating is also a complicated function of load impedance, getting worse for unloaded conditions. No device capable of anticipating its amount has been very satisfactory.

The *Z1492* cathode, running normally at 560 watts, would have its life shortened by a factor of 7 by the addition of 155 watts of back heating, so it is worth while to compensate for it.

A regulating circuit, shown in Figure 11, compares the resistance of the filament with a reference resistor. As the filament temperature rises, its resistance increases. This generates an error signal in the servo circuit which reduces the heating current. The circuit as shown will compensate for three-fourths of the back-heating power in a *Z1492* magnetron. More elaborate controllers with better regulation have been designed, but the circuit shown has the advantage of simplicity and of failing safe. If a tube loses conductivity, the magnetron filament becomes colder.

Radio-Frequency Transmission Lines. The *Z1492* is designed to feed into a $3^1/_8$-inch coaxial transmission line of 53 ohms characteristic impedance. Lines with dielectric beads supporting the center conductor are not desirable, due to reflections from the beads and dielectric losses in them. Stub-supported lines, having an inner conductor of $1^1/_4$-inch diameter and an outer conductor of 3.027-inch inside diameter, have been used and have been found satisfactory.

Applications of 5-Kw Magnetron

Experimental work on dielectric heating at 915 megacycles has been carried on by

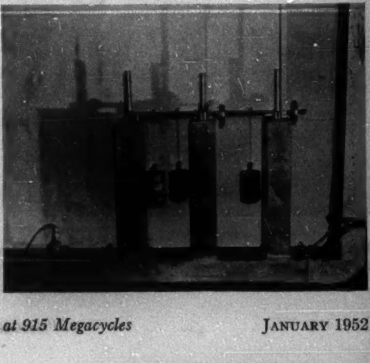


Figure 17. Wave guide test equipment for 1,000-megacycle tubes. Left to right are a standing wave detector, triple stub impedance matcher, and vertical water load

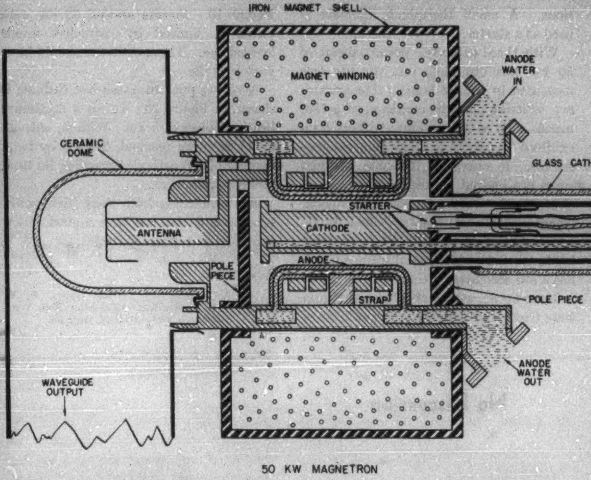


Figure 18. Mounting of high-power magnetron. The cathode shown is a secondary-emitter

several departments of the General Electric Company and also by other organizations.

To facilitate experiments, a number of 5-kw heater units have been built and sold. This heater unit, shown in Figure 12, contains a *Z1492* magnetron, power supplies, filament temperature regulator, triple stub radio-frequency impedance matcher, and directional couplers to read directly the power in the transmitted and reflected waves. In use, matching the external load to the oscillator is done by adjusting the two controls on the impedance matcher until the reflected power meter reads zero.

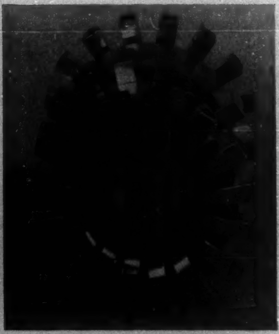


Figure 19. Anode structure of high-power tube

In microwave heating, even more than low-frequency heating, each job presents a new problem in coupling power to the load. One example which has been studied is the heating of precooked frozen dinners. The heater shown in Figure 13 is a specialized unit for this purpose, developed by P. W. Morse and H. E. Revercomb.[1] It contains a resonant cavity, inside which the food is rotated on a phonograph-like turntable to average out the standing wave pattern.

Another example of coupling methods is in the preheating of thermosetting plastics. Here the pieces are small and uniform, but a very uniform temperature is required. One method which has been used is the resonant cavity shown in Figure 14. The transmission line is coupled by an inductive loop to the cavity, which is tuned to resonance by moving one side. The plastic preform pellet is placed at the center of the cavity in the region of most uniform field. A support of low-loss dielectric prevents burning at the points of contact with the metal walls.

In this cavity, fed with 915-megacycle power, phenolic plastics can be preheated to a molding temperature of 190 degrees centigrade in one second.

Figure 15 shows an experimental plastic preheater developed by R. L. Borchert of the General Electric Plastics Division. In this machine, preform pellets are fed through a resonant cavity on an endless belt, emerging at molding temperature. Quarter-wave chokes reduce radiation through the entrance and exit holes in the cavity.

High Power Magnetron

Experimental magnetrons have been built to give up to 50-kw output at 1,000 megacycles. Figure 16 shows a cut-open view of a tube which was developed as part of a Signal Corps contract. Similar tubes may be built to a degree of ruggedness and reliability suitable for industrial dielectric heating.

This magnetron operates at about 14 kv, with a magnetic field of 1,500 gausses. Plate efficiency is from 50 per cent to 60 per cent.

At the high-power level it is desirable to use wave guides for transmission lines. Figure 17 shows the 4 by 8 inch waveguide plumbing used for testing tubes. The magnetron output is designed to couple directly into the wave guide, as shown in Figure 18, by a quarter-wave antenna penetrating the guide. Surrounding the antenna is a ceramic dome which forms the vacuum seal. The ceramic material, a high-alumina vitreous

Figure 20. Tungsten spiral cathode. [Filament wires are 0.100 inch in diameter

body known as Aluminite, will stand much higher powers than glass.

The anode structure of this large tube, shown in Figure 19, consists of 16 anode $3^{1}/_{2}$-inches long by 1/4 inch wide, each carrying cooling water. It will dissipate 80 kw.

The limit on power in these continuous-wave magnetrons is in the cathode, where back heating is severe. Experiments were carried on with water-cooled cathodes coated with good secondary-emitting materials. Such a cathode is shown in the tube in Figure 18. With a secondary electron yield ratio of 3 to 1, the back-bombarding electrons generate enough secondary emission to sustain oscillations. A small thermionic cathode is used as a starter.

With these cathodes, power output of 50 kw was obtained, but the life was measured in tens of hours. The secondary emitting surfaces consisted of active metals such as magnesium and beryllium oxidized to a depth of a few hundred angstroms. In operating the tube, residual gases are ionized by the high density of electrons between cathode and anode. These positive ions bombard the cathode and are believed to sputter off the oxide layer, leaving a clean metal surface which has a low secondary yield.

Thermionic cathodes coated with thoria were tried—one is in the cut-open tube of Figure 16. Powers around 30 kw were obtained, limited by cathode-to-anode breakdown. These cathodes should have very long life.

A strong pure tungsten hot cathode is shown in Figure 20. It has a maximum heating power of 5 kw. With it, outputs of 35 kw were reached, limited by back heating of the cathode exceeding its total power requirement.

The conclusion from these high power tube tests is that we have a good 30-kw oscillator at 1,000 megacycles.

Reference

1. UHF HEATING OF FROZEN FOODS, P. W Morse, H. E. Revercomb. *Electronics* (New York, N. Y.), volume 20, October 1947, page 85.

No Discussion

Characteristic Impedance of Rectangular Coaxial Transmission Lines

YOUNES A. OMAR
ASSOCIATE MEMBER AIEE

C. FRANK MILLER
MEMBER AIEE

EXACT mathematical relations between the characteristic impedance and the physical dimensions of circular concentric and eccentric coaxial lines, as well as of coaxial lines having square outer and circular inner conductors[1] have been developed and experimentally verified. Messinger and Higgins[2] were able to develop formulas for the determination of the inductance of coaxial busses composed of square tubular conductors, at power frequencies. Their calculations were based on finding the self and mutual geometric mean distances of the conductor cross sections.

An exact mathematical treatment of coaxial lines composed of rectangular inner and outer conductors is exceedingly difficult, if not impossible, because of the discontinuities at the conductor corners.

The purpose of this research work was to determine experimentally the characteristic impedance of several coaxial rectangular transmission lines, and to establish from the results thus obtained an empirical relation between the line dimensions and its characteristic impedance.

The standing-wave measurement method was found to be most suited for this work. If a line of unknown characteristic impedance is terminated by a known impedance, one can find, by calculation or by the use of the proper impedance charts, the unknown characteristic impedance of the line. The measurements were made at approximately 430 megacycles per second.

Characteristic Impedance of Coaxial Lines

For a low-loss radio-frequency transmission line[3,4] the characteristic impedance is a pure resistance and practically independent of frequency, having the value

$$Z_0=(L/C)^{1} \qquad (1)$$

and the velocity of phase propagation is defined by

$$V=\lambda f=1/(LC)^{1/2} \qquad (2)$$

where

Z_0 = line characteristic impedance in ohms
L = series inductance per-unit length of the line in henrys
C = shunt capacitance per unit length of the line in farads
V = velocity of phase propagation
f = frequency
λ = wave length

It is a property of the electromagnetic field that, for a dissipationless line in free space, L and C are so related that their product is constant, and that the phase velocity is equal to the velocity of light. Based on these statements, the inductance and capacitance per unit length of a transmission line can be computed from a knowledge of the characteristic impedance and vice versa.

CIRCULAR COAXIAL LINES

The characteristic impedance of circular coaxial lines[5] is given by

$$Z_0=138 \log_{10} A/a \text{ ohms} \qquad (3)$$

where

a = outside radius of the inner conductor
A = inside radius of the outer conductor

Curve I of Figure 1 shows the graphical representation of equation 3.

SQUARE COAXIAL LINES

Messenger and Higgins[2] have derived a formula for the a-c inductance of square tubular lines at frequencies up to several hundred cycles. Figure 2 shows a plot of the equation derived by these authors. This equation was based on the assumption of infinitely thin conductors, and thus it can be applied to very high frequencies where the field is confined between the outer surface of the inner conductor and the inner surface of the outer conductor The characteristic impedance of such square coaxial lines, shown in curve II of Figure 1, was computed using relations 1 and 2 for values of L taken from Figure 2.

CIRCULAR CONDUCTORS IN SQUARE TROUGHS

Sidney Frankel,[1] employing the method of conformal transformation and the method of images, jointly, was able to deduce a formula for the characteristic impedance of concentric lines having circular inner conductors and square outer ones:

$$Z_0=138 \log_{10} 1.079\, A/a \text{ ohms} \qquad (4)$$

where $2A$ is the length of the side of the outer conductor, and a is the radius of the inside circular conductor. Curve III of Figure 1 shows a plot of this equation.

CONCENTRIC CONFOCAL ELLIPTIC TRANSMISSION LINES

The problem of the capacitance of the confocal elliptical capacitor[6] has been solved, and the capacitance per unit length of such a capacitor is known. Using relations 1 and 2 one can find the characteristic impedance of such configuration to be:

$$Z_0=138 \log_{10} \frac{A/a+\{(A/a)^2-(d/a)^2\}^{1/2}}{1+\{1-(d/a)^2\}^{1/2}} \qquad (5)$$

where

Paper 52-75, recommended by the AIEE Special Communications Applications Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21-25, 1952. Manuscript submitted August 17, 1951; made available for printing December 14, 1951.

YOUNES A. OMAR is with the Ports and Lighthouses Administration, Alexandria, Egypt, and C. FRANK MILLER is at The Johns Hopkins University, Baltimore, Md.

The authors wish to express their appreciation to Dr. Jean V. Lebacqz for valuable suggestions and criticisms.

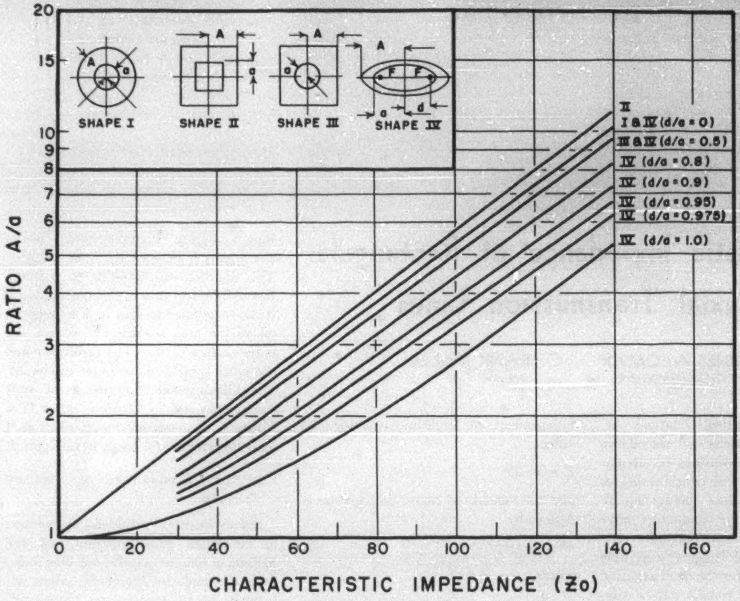


Figure 1. Characteristic impedance of circular, square, square outer - circular, inner and elliptic coaxial lines

A = semimajor axis of the exterior ellipse
a = semimajor axis of the interior ellipse
d = distance between the center point of the capacitor and the focii

If B be the semiminor axis of the exterior ellipse and b be the semiminor axis of the interior ellipse then for a confocal ellipse we have the relation

$$d^2=A^2-B^2=a^2-b^2 \tag{6}$$

If in equation 5 the semiminor axis b of the inside ellipse is very small compared to a, then in this case d is very closely equal to a and the equation becomes:

$$Z_0=138\log_{10}\left\{A/a+[(A/a)^2-1]^{1/2}\right\} \text{ ohms} \tag{7}$$

Also, if in equation 5, d is made equal to zero, which is the case of a circular coaxial line, the equation reduces to equation 3. Figure 1 shows a plot of equation 5 for different values of d/a.

STANDING WAVES ON TRANSMISSION LINES

If a transmission line is terminated with a load different from its characteristic impedance, energy is reflected back into the line and the combination of the incident, and reflected signals produces a standing wave. The voltage standing wave ratio is usually defined as:

$$\rho=\left|\frac{E_{\max}}{E_{\min}}\right| \tag{8}$$

however, in this paper the voltage standing wave ratio (*VSWR*) is taken as $1/\rho$. From a knowledge of the *VSWR* and the positions of voltage minima along a transmission line, one can easily determine the nature of the load impedance which terminates the line.[3-4]

Description of Equipment and Apparatus Used

Although the purpose of this paper is to measure the characteristic impedance of rectangular coaxial lines, it was found necessary to build some transmission lines of known characteristic impedances. This was considered to be rather important since such lines could be used for checking the method of measurement, making necessary adjustments and measuring the impedance and characteristics of certain types of terminations.

The coaxial lines of known characteristic impedances which were used in the experimental work reported in this paper were of two types:

1. Circular coaxial.
2. Circular inner and square outer transmission lines.

The characteristic impedance of such lines can be computed, and these lines are easy to build.

THE DESIGN OF CIRCULAR COAXIAL SLOTTED TRANSMISSION LINE

A circular coaxial line was built especially for this experimental work. The line itself is 10 feet long and is composed of a copper inner conductor of $^5/_8$-inch outside diameter, and a copper outer conductor having $1^1/_2$-inch inside diameter. A 3-foot section of this line was slotted to accept a traveling probe for the measurement of the voltage standing wave ratio. The carriage, which accommodates the probe, is fitted with a gear and pinion, and moves smoothly along two parallel lengths of gear rack, one on each side of the slot. The two conductors, of which the line is composed, were made concentric by using thin circular washers made of Polystyrene and spaced about 1 foot apart.

THE DESIGN OF RECTANGULAR COAXIAL SLOTTED TRANSMISSION LINE

Since the problem involves the measurement of the characteristic impedance of several rectangular lines, it was more convenient and economical to construct a built-up line, that is, the outer conductor

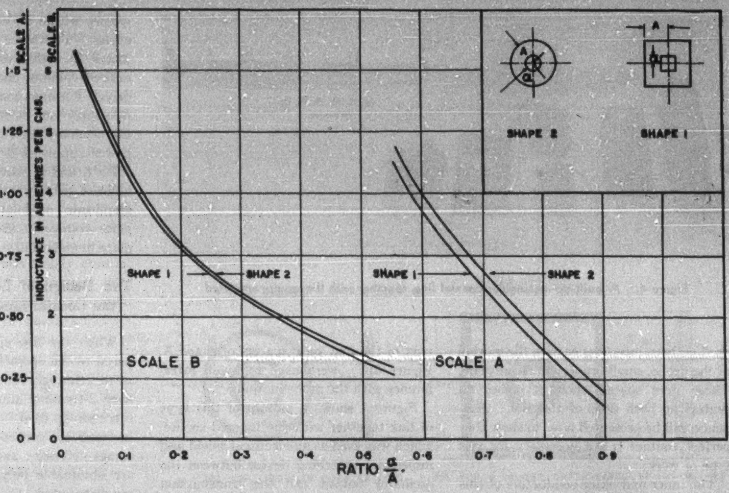


Figure 2. Messinger and Higgins formulas for inductance of square coaxial busses

of which was built of four flat brass bars joined together to give a hollow rectangular section of the required dimensions. Figure 3 shows a detailed drawing of this line. The inner conductor could be made circular or rectangular, and the outer conductor had upper and lower plates, having a cross section $2^1/_2$ inches by 1/8 inch. The line was about $9^3/_4$ feet long. A slot 1/8-inch wide and 3 feet long was cut in the top plate to accommodate a traveling probe. The two vertical sides of the outer conductor were made of flat brass bars 1/4-inch thick and of different heights.

The width of the outer conductor could be changed from about 3/4-inch to about $2^1/_4$ inches, and the height from about 1/2 inch to any value required. Beside changing the dimensions of the outer conductor, the inner conductor could be changed to any rectangular section, a circular rod, square bar, or any other configuration.

The four plates composing the outer conductor were held together by C-clamps spaced about 4 inches apart to insure good electrical contact. The clamps were applied in such a way that clamping pressure was principally on the vertical sides, and as close as possible to the inner face of these sides. In the slotted part of the line, since only a few clamps could be

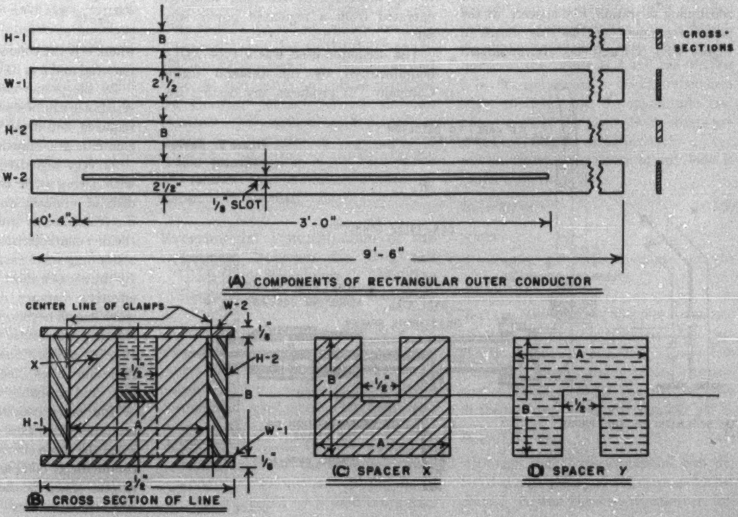

H-1
W-1
H-2
W-2
B
2½"
B
2½"
⅛" SLOT
CROSS-SECTIONS
0'-4"
3'-0"
9'-6"
(A) COMPONENTS OF RECTANGULAR OUTER CONDUCTOR
CENTER LINE OF CLAMPS
W-2
⅛"
X
½"
H-2
B
H-1
A
W-1
⅛"
2½"
(B) CROSS SECTION OF LINE
B
½"
A
(C) SPACER X
A
B
½"
(D) SPACER Y

Figure 3. A built-up rectangular line

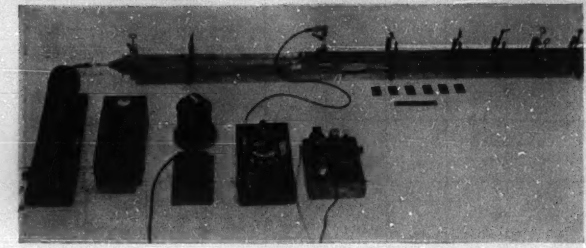


Figure 4. A built-up rectangular coaxial line together with the equipment used

used without interference with the motion of the probe, small brass rivets were used. These were spaced about 4 inches on centers on both sides of the slot. Evidence will be presented later to show that perfect contact is not necessary for this type of work.

The inner and outer conductors of this line were made coaxial, along its entire length, by means of 1/16-inch thick Polystyrene spacers arranged as shown in Figure 3 (C) and (D). The inside and outside dimensions of these spacers corresponded to those of the inner and outer conductors of the line. A different set of spacers was required for every line. Ten spacers, spaced 1 foot apart, were found to be sufficient to insure rigidity of the line. Instead of machining rectangular holes in the spacers to fit the inner conductor, which is a rather difficult and time consuming operation, two spacers, at the point of support, are used as shown in Figure 3(B), (C), and (D) In the slotted part of the line, only spacers of shape A, Figure 3(C), were used to avoid interference with the probe motion.

Figure 4 shows a picture of this type of line together with the tapered section which was used as an electrical shield and impedance matching device between the oscillator output and the transmission line.

The Probe, Oscillator, and Detector Amplifier

Figure 5 shows the details of the probe which was designed for use at frequencies in the neighborhood of 400 megacycles. It was built from standard connectors with an arrangement to permit variation of the depth of penetration of the probe in the coaxial line.

The indicator amplifier unit was similar to one described by Kallman[7] and was operated from a regulated power supply.[8]

The oscillator used was a type *757-A* manufactured by the General Radio Company. It produces any desired frequency within the range 150 to 600 megacycles. The wave length could be adjusted to within 0.01 centimeter of the desired value by means of a slow-motion drive. For most conditions of loading, the frequency can be relied on to be within the accuracy limit of 2.5 per cent. The power supply is a General Radio type *757-P1*, and is provided with a meter to indicate grid current. The oscillator was amplitude modulated by varying the plate voltage in the normal fashion for plate modulation.

The Design of Loads to Terminate the Coaxial Lines

When the line spacing is small compared to the operating wave length, that is, at relatively low frequencies, an ordinary 2-terminal lumped resistance can be successfully used. At high frequencies, however, the concept of guided waves becomes necessary and satisfactory results are obtainable only if the terminations are distributed.

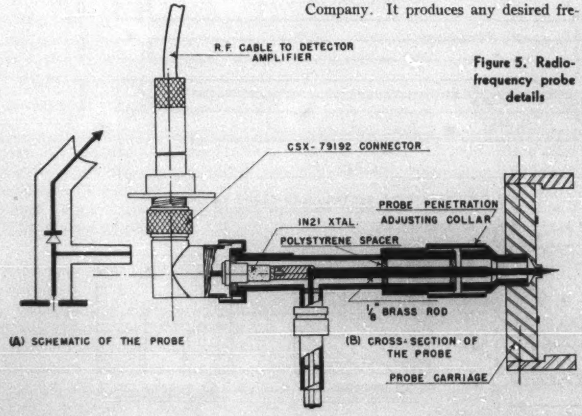
Ideally this is done for the coaxial line by closing the end of the line by a disk resistor as shown in Figure 6. In practice these distributed terminations are usually made by spraying some form of carbon on a dielectric support such as bakelized paper. Under certain conditions, a piece of carbon coated paper serves as a perfect resistance termination.[9] Referring to Figure 6, the resistance of the disk between the two conductors is given by

$$R = \frac{1}{2\pi\sigma h}\log_e (A/a) \qquad (9)$$

where σ is the conductivity of the coating material and h is its thickness.

By the proper choice of σ and h, this resistance can be made equal to any value required, and it can be made equal to the characteristic impedance of the line.

At very high frequencies, experiments with a concentric line[10] terminated by a disk of uniform resistive material (the d-c resistance of which is made equal to the line characteristic impedance) showed a standing wave ratio different from one. By the use of a short-circuiting plate at an effective distance of one quarter of a wave length beyond the graphite film, the standing wave ratio was found to be very close to unity. The explanation is that radiation takes place from the graphite film, and probably from a small length of the inner conductor just inside the graphite film. This effect is only marked when the mean circumference of the outer conductor is comparable with the wave length, which is the condition that the disk shall radiate appreciable energy.



R.F. CABLE TO DETECTOR AMPLIFIER
CSX-79192 CONNECTOR
PROBE PENETRATION ADJUSTING COLLAR
IN21 XTAL.
POLYSTYRENE SPACER
1/8" BRASS ROD
(A) SCHEMATIC OF THE PROBE
(B) CROSS-SECTION OF THE PROBE
PROBE CARRIAGE
SCALE 1:1

Figure 5. Radio-frequency probe details

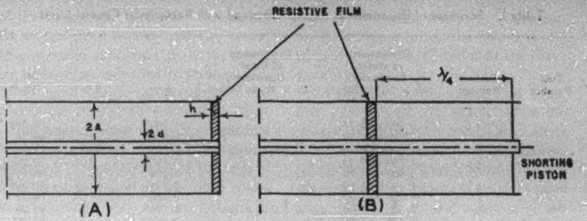
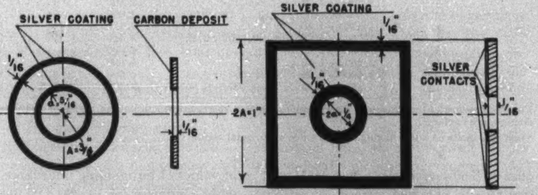


Figure 6. Resistive disk termination



(A) FOR CIRCULAR COAXIAL LINE (B) FOR A SQUARE OUTER-CIRCULAR INNER LINE

Figure 7. Two different shapes of terminations

At this wave length, the addition of the quarter-wave short-circuiting plate serves to screen the disk, thus prohibiting the radiation and making the line termination correct.

In the problem at hand, we are not actually interested in terminating a line in its characteristic impedance, but in terminating the line by a resistive disk whose effective resistance, at the operating frequency, is more or less equal to its d-c resistance or bearing a certain relation to it. Since the experimental work involved in this paper was mostly done at one frequency (430 megacycles per second), the line dimensions are small compared to the wave length and negligible energy is radiated from the back of these disks, and no short-circuiting plate is required at an effective distance of one-quarter of a wave length behind these washers. This statement was proved experimentally as will be shown later.

Choice of the resistive coating

Graphite, carbon, and metallic coating were tried in this work. Washers made of 1/16-inch Polystyrene and other low-loss dielectrics were cut having the shape of the transmission line, as shown in Figure 7. The outer circumference was made to fit the inside of the outer conductor, and the inner to fit the outside of the inner conductor. The electrical contact between the termination and the line was accomplished by using a 34 per cent silver (Number 4817), manufactured by the DuPont Company. The contact surfaces, as well as about 1/16 inch of the surface of these discs were painted with this silver solution and baked for about 10 hours at 100 degrees centigrade.

The particular deposit used to build all of the line terminations employed in this work was type *3051* Rescon made by the General Electric Company. It is a low-resistance surface carbon deposit, which when baked has a resistivity of the order of 1,000 ohms per square. The resistance depends on the baking time which ranges from 15 hours at 100 degrees centigrade to 7 hours at 130 degrees centigrade to 3 hours at 150 degrees centigrade.

Graphite was also used, and experimental results showed that the effective resistance of several samples does not change with frequency up to about 500 megacycles per second, and that it differs from the d-c values by less than about 2 per cent. But, due to the high temperature coefficient of graphite, the resistance was found to be rather unstable and changed pronouncedly with temperature and aging. For this reason deposited carbon resistors were preferred.

Samples were also made by evaporating aluminum, under vacuum, on Polystyrene washers to a thickness of about 10^{-8} centimeter. Contact between the aluminum film and the transmission line conductors was accomplished with the silver paint. Experiments with these samples proved them to be much superior to the deposited carbon samples. Their effective resistances were found to be very close to their d-c values (within less than 1 per cent), but due to the different coefficient of expansion of aluminum and silver, it was rather difficult to maintain the electrical contact between the two elements. This method, beside being very costly, does not offer an easy means of controlling the value of the resistances, and for these reasons aluminum coated resistors were not used.

Experimental Determination of the Relation Between the D-C and Ultra-High-Frequency Effective Resistance of the Disc Resistors

Tests were made with a circular coaxial line and a line having a square outer and circular inner conductors to measure the impedance of 30 different termination resistances. All tests were made at the same frequency (430 megacycles per second), and the method of measurement was the standing wave method. In this method, the far end of the line is shorted by a solid brass disc, and a point of minimum voltage is located as accurately as possible near the center of the slotted part of the line. This minimum point is taken as the reference point. The load is made to replace the short-circuiting disk and the shift θ in the position of the minimum point is read from a scale, and the ratio of the voltage minimum to voltage maximum is obtained from the probe meter. This ratio is corrected[11] and the corrected ratio is called throughout this paper, the Voltage Standing Wave Ratio, or *VSWR*. Having determined the values of θ and the *VSWR* it is then convenient to use an impedance chart to determine the magnitude and phase angle of the load. Such charts are given by Meagher and Markly.[11]

The effective resistance of any load is found from the relation

$$R_{\text{eff}} = Z_1 . \cos \theta_1 \qquad (10)$$

where

Z_1 = terminating load impedance, at the operating frequency
θ_1 = load phase angle

Tests showed that the effective resistance of these terminations, using this particular carbon Rescon, was always less than their direct-current values irrespective of the shape of the line and the ratio of the two resistances was found to be:

$$R_{\text{eff}} = 0.793\ R_{\text{dc}} \qquad (11)$$

Although the above expression was derived at a frequency of 430 megacycles per second, it was checked at another fre-

Table I. Summary of Experimental Results Obtained with Rectangular Coaxial Lines

Test Number	Section Number	Dimensions of Outer Conductor, (2A × 2B) in Inches	Dimensions of Inner Conductor, (2a × 2b) in Inches	A/a	B/b	Z_0, Ohms
	1	0.892 by 0.75		1.784		56.70‡
	2	1.250 by 0.75		2.500		68.10
A	3	1.500 by 0.75	0.500 by 0.125	3.000	6.0	75.00
	4	1.750 by 0.75		3.500		80.00
	5	2.000 by 0.75		4.000		84.60
	1	1.110 by 1.00		2.220		70.50‡
	2	1.250 by 1.00		2.500		74.20
B	3	1.500 by 1.00	0.500 by 0.125	3.000	8.0	80.36
	4	1.750 by 1.00		3.500		87.00
	5	2.000 by 1.00		4.000		91.10
	1	1.340 by 1.25		2.680		83.50‡
C	2	1.500 by 1.25	0.500 by 0.125	3.000	10	87.14
	3	1.750 by 1.25		3.500		93.00
	4	2.000 by 1.25		4.000		97.70
	1	1.576 by 1.50		3.152		92.00‡
D	2	1.750 by 1.50	0.500 by 0.125	3.500	12	96.30
	3	2.000 by 1.50		4.000		102.5
	4	2.250 by 1.50		4.500		106.4
F	1	0.900 by 0.40	0.500 by 0.125	1.800	3.2	42.04
E*	1	1.250 by 1.25	0.500 by 0.500	2.500	2.5	53.10

‡The dimensions of these sections were chosen such that the inscribed ellipses, to the inner and outer conductors, are confocal.

*Test E is made on a line having a wave guide as its outer conductor.

quency (350 megacycles per second), and the results obtained agreed to within better than 2 per cent. Expression 11 can thus be used to determine the characteristic impedance of any coaxial line, using the same type of line terminations, if the d-c resistance (R_{dc}) and the ratio Z_1/Z_0 or the ratio R_{eff}/Z_0 are known. In this work, impedance charts,[11] were used and the characteristic impedance of the line was given by:

$$Z_0=\frac{R_{dc}\times 0.973}{\cos\theta_1}\cdot\frac{Z_0}{Z_1} \qquad (12)$$

The tests also showed that these disk terminations are capacitive. This is partly due to the self-capacitance of the resistors and partly due to their mountings. When examined under the microscope, the surface appears rough and mountainous. The self-capacitance of the resistors is believed to be due to the fact that the carbon deposit particles are rather spherical in shape. At the points of contact between these particles the contact resistance is higher than that of the particles themselves. Adding to that, the capacitances between contact points, one gets a combination whose effective resistance is slightly less than the d-c resistance and whose impedance is capacitive.

Relation 12 is thus an inherent property of the particular carbon solution used, independent of the shape of the resistor. Having established this experimental relation, one is in a position to find the characteristic impedance of different shapes and combinations of coaxial lines.

MEASUREMENTS ON RECTANGULAR COAXIAL LINES

Tests were made on twenty different rectangular coaxial lines, with a variety of values of terminating impedance. The outer conductor for 19 of the 20 lines was built up of four separate elements, clamped together as described before. Great care was taken to insure good electrical contact between the four members of the outer conductor by increasing the number of C-clamps. It was felt, however, that the question of good electrical contact should be investigated experimentally to remove any doubt concerning the results to be obtained.

By using C-clamps 4 inches apart along the line and on both sides, one can argue about the continuity of the electrical contact along the line. In the worst case, one can assume that there is no electrical contact whatsoever between the four members of the outer conductor except, perhaps, at the two terminals of the line where the oscillator and the disc loads are connected. This condition can be realized in practice by insulating the four members from each other, except at the two terminals, by means of cellulose acetate insulating adhesive tape.

Measurements of the *VSWR* and phase angle with and without the insulating tape were taken for ten different disk terminations. From these results the ratio Z_1/Z_0 and the loads phase angles (θ_1) were obtained. The results obtained with the tape gave values about 4 per cent higher than those obtained without the tape. This difference between the two sets of readings is actually the accuracy of this method of measurement. From such test one can conclude that the method of clamping used during this work is adequate as far as standing wave ratios measurement is concerned.

In all the following tests the inside conductor was kept the same while the height of the outer conductor was changed from about 3/4 of an inch to $1^1/_2$ inches in steps of 1/4 inch and its width from about 3/4 inch to $2^1/_4$ inches, also in steps of 1/4 inch. In this way it was possible to investigate the effect, on the line characteristic impedance, of changing one parameter at a time.

All tests were made at the same frequency (430 megacycles per second), same probe insertion, and same silicon crystal detector.

One of the rectangular coaxial lines used in this work was made of a piece of 3-centimeter wave guide, as the outer conductor, while the inner conductor is a length of a solid rectangular bar. This section was chosen as an additional check on the built-up line, as will be discussed later. Table I gives a summary of the data for the line sections studied, as found experimentally.

Analysis and Discussion of Results

GENERAL CONSIDERATIONS

One might notice from Table I that throughout every test the dimensions *a*, *b*, and *B* were kept the same while the width of the outer conductor *A* was the only variable. Thus from each one test it is possible to obtain an idea of the characteristic impedance of a rectangular coaxial line as function of the dimension *A*. However, the value of *B* was different in each test thus providing another possible relationship between the line characteristic impedance and the dimension *B*. In all tests the dimensions *a* and *b* were kept the same, that is, only one inside conductor was used, except for the square coaxial line of test *E*.

CONFOCAL ELLIPSES

Conductors of rectangular cross section are very difficult to treat mathematically. The exact solution of the potential problem for a rectangular cross section can be solved by making use of the Schwartz Transformation, but this transformation can hardly be used to handle a problem such as a rectangular coaxial line. For this reason we must be satisfied with approximating such configuration with another one which can be handled mathematically. The best approximation of a rectangular cross section is an elliptic section, and the best approximation of a coaxial rectangular section is a pair of coaxial ellipses. But it is very difficult to find, for example, the capacitance of a capacitor composed of two elliptic shells unless these two ellipses are confocal, that is, have common focii.

Keeping this idea in mind, the first

rectangular coaxial lines of tests *A*, *B*, *C* and *D* were chosen such that the inscribed ellipses to the inner and outer capacitor are confocal. These four sections allow us to examine the relation between the characteristic impedances of our sections as measured experimentally, and those of the inscribed confocal ellipses as computed mathematically.

The characteristic impedance of an air-filled confocal ellipse section was given in equation 5 together with the condition imposed on the semiminor and semimajor axes of both the inner and outer ellipses in order for the ellipses to be confocal, that is, equation 6. From this relation we can readily see that for every inside ellipse, that is, for fixed a and b, there are an infinite number of outer ellipses which will satisfy equation 6, by choosing the proper values for the dimensions A and B. But, once we fix a value, say for dimension B, then there is only one outer ellipse which is confocal with the inside one. This last case corresponds to our problem since we used only one inside conductor for most of the tests, and for every one test, we kept the dimension B unchanged.

Assuming that the four particular sections, mentioned before, are confocal ellipses instead of being rectangular, we can use equation 5 to find their characteristic impedances. This assumption is actually equivalent to neglecting the sharp edges of the rectangular sections. Comparing these calculated values and the actual characteristic impedance of the rectangular sections as measured experimentally, it was found that the latter values are less than those of the confocal ellipses. This result was expected since by neglecting the edges we have decreased the line inductance. Since the ratio of the area of a rectangle to that of the inscribed ellipse is a constant $(4/\pi)$, we are justified in taking an average value for the characteristic impedance of a coaxial rectangular line whose inscribed ellipses are confocal. This assumption will be justified when we consider the characteristic impedance of a square coaxial line, later in this section.

From this experiment, it was found that the characteristic impedance of such lines is that of the corresponding confocal ellipse multiplied by 0.971. Thus the characteristic impedance of a rectangular coaxial line whose inscribed ellipses are confocal, may be given by the following equation:

$$Z_0 = 134 \log_{10} \frac{A/a + \{(A/a)^2 - (d/a)^2\}^{1/2}}{1 + \{1 - (d/a)^2\}^{1/2}} \tag{13}$$

where $d^2 = A^2 - B^2 = a^2 - b^2$.

Figure 8 shows a plot of this equation for different values of d/a including the value 0.968 which was fixed by the dimensions of the inside conductor. One can see from Figure 8 that the four first sections of tests *A*, *B*, *C*, and *D* fit the curve $d/a = 0.968$.

To support the statement that equation 13 represents the characteristic impedance of rectangular coaxial lines whose inscribed ellipses are confocal, we may refer to curve II of Figure 1 which gives the characteristic impedance of coaxial square lines. The approximate equation of this curve, as found from curve II, is:

$$Z_0 = 134 \log_{10} A/a \tag{14}$$

This equation is actually the upper limit of equation 5, one which has the ratio $d/a = 0$. Also the experimental result obtained in test *E* checks very closely the above equation. This justifies our assumption that the percentage decrease in the characteristic impedance of rectangular coaxial lines, whose inscribed ellipses are confocal, due to sharp corners is independent of the line dimensions.

THE DEVIATION FROM THE CONFOCAL ELLIPSE SECTION

Plotting the experimental results given by Table I, on Figure 8, one can see at once, for every section, that by increasing the dimension A of the outer conductor (while keeping all other dimensions unchanged), the characteristic impedance of the section increases following a semi-

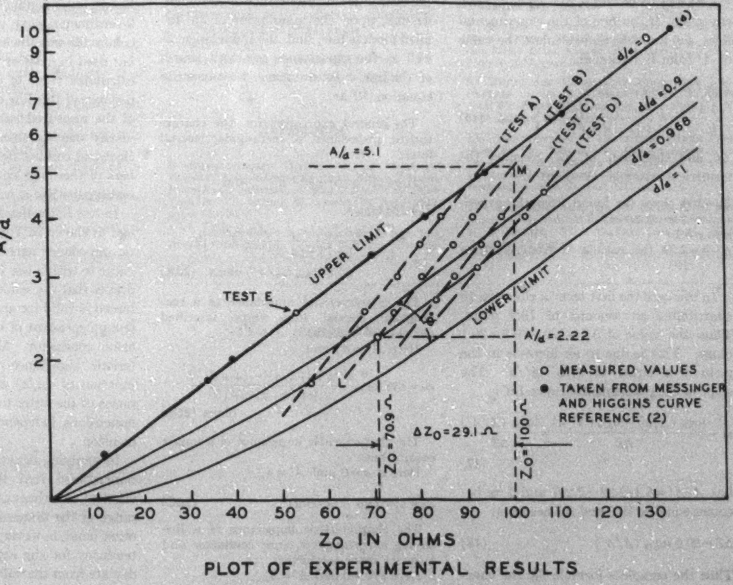


Figure 8. Plot of experimental results

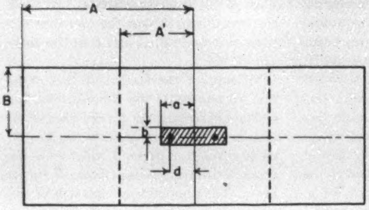


Figure 9. Relations between the different dimensions

logarithmic law. For the four different tests, A, B, C, and D, we get four straight lines on semilogarithmic paper. These lines were found to be very nearly parallel, thus we are in a position to find an imperical formula for the characteristic impedance of any rectangular coaxial line.

To do that we shall define a dimension A' such as:

$$A'=(B^2+a^2-b^2)^{1/2} \tag{15}$$

and is the width of the outer conductor which is necessary to make the inscribed ellipses (keeping all other dimensions B, a and b constant) confocal. Thus for sections whose width A was taken such that the inscribed ellipses are confocal A' is equal to A. For such sections then equation 5 gives their characteristic impedance directly. For other sections in which A' is not equal to A, one must add a correction term to equation 5.

One way to do this is to take any point, say point M, on one of the experimental lines, say line LK, and calculate the value of A' from Equation as:

$$Z_0=134\log_{10}\times \left\{\frac{A'/a+[(A'/a)^2-(d/a)^2]^{1/2}}{1+[1-(d/a)^2]^{1/2}}\right\}+\Delta Z \tag{16}$$

To find the value of ΔZ, consider the numerical example: At point M

$Z_0=100$ ohms (as found from the experimental curve LK)

$A/a=5.1$

$A'/a=2.22$ (as calculated from equation 15)

In this case the first term of equation 16 contributes an amount of 70.9 ohms. Thus the value of $\Delta Z=100-70.9=29.1$ ohms. This is due to an increase in the ratio A/a equal to $A/a-A'/a$. The slope of the experimental curve LK is:

$$m=\frac{\log_{10}(A/a)-\log_{10}(A'/a)}{\Delta Z}=\frac{\log_{10}(A/A')}{\Delta Z} \tag{17}$$

but $A/A'=5.1/2.22=2.297$ and $1/m$ becomes equal to 80.6 and ΔZ becomes:

$$\Delta Z=80.6\log_{10}(A/A') \tag{18}$$

Thus the complete formula for the characteristic impedance of any coaxial rectangular transmission line is:

$$Z_0=134\log_{10}\frac{\{(A'/a)^2-(d/a)^2\}^{1/2}+A'/a}{\{1-(d/a)^2\}^{1/2}+1}+80.6\log_{10}(A/A')\text{ ohms} \tag{19}$$

where

A' is given by equation 15

$2A$ = inner width of the outer conductor

$2B$ = inner depth of the outer conductor

$2a$ = outer width of the inside conductor

$2b$ = thickness of the inside conductor

Figure 9 shows these different dimensions.

Equation 19, in summary, means: To obtain the characteristic impedance of a rectangular coaxial line whose dimensions A, B, a, and b are known, we use equation 15 to find a dimension A' which upon replacing dimension A will give two confocal ellipses inscribed to the coaxial rectangular conductor sections. Knowing the characteristic impedance of such lines, we can, upon the assumption of an air-filled lossless line, find the inductance as well as the capacitance per unit length of the line. In summary, we can write equation 19 as:

The general expression for the characteristic impedance of rectangular coaxial lines:

Here $A'=(B^2+a^2-b^2)^{1/2}$, and

$$Z_0=134\log_{10}\times\left[\frac{A'/a+\{(A'/a)^2-(d/a)^2\}^{1/2}}{1+\{1-(d/a)^2\}^{1/2}}\right]+80.6\log_{10}(A/A')\text{ ohms} \tag{20A}$$

The characteristic impedance of a rectangular coaxial line whose inscribed ellipses are confocal:

Here $A'=A$, and

$$Z_0=134\log_{10}\frac{A/a+\{(A/a)^2-(d/a)^2\}^{1/2}}{1+\{1-(d/a)^2\}^{1/2}}\text{ ohms} \tag{20B}$$

The characteristic impedance of a square coaxial line:

Here $d/a=0$ and $A'=A$,

$$Z_0=134\log_{10}A/a\text{ ohms} \tag{20C}$$

The characteristic impedance of a line having a rectangular outer conductor and a square inner conductor:

Here $d/a=0$ and $A'=B$,

$$Z_0=134\log_{10}(A'/a)+80.6\log_{10}(A/A')\text{ ohms} \tag{20D}$$

The characteristic impedance of a line having a rectangular outer conductor and a ribbon inner conductor:

Here $d/a=1$ and $A'=(B^2+a^2)^{1/2}$,

$$Z_0=134\log_{10}[A'/a+\{(A'/a)^2-1\}^{1/2}]+80.6\log_{10}(A/A')\text{ ohms} \tag{20E}$$

The characteristic impedance of a line having a rectangular outer conductor and a ribbon inner conductor when the inscribed ellipses are confocal:

Here $d/a=1$ and $A'=A$,

$$Z_0=134\log_{10}[A/a+\{(A/a)^2-1\}^{1/2}]\text{ ohms} \tag{20F}$$

Equations 20 (A-F) will give the characteristic impedance of any rectangular coaxial line, rectangular outer conductor square inner conductor line, square coaxial line, and rectangular outer conductor ribbon inside conductor line.

To test the "fit" of these formulas, we shall go backward and apply them to all the different sections used in this experimental work and compare the calculated with the measured values, as shown in Table II. From Table II, one can see that the calculated values of the characteristic impedance of rectangular coaxial lines, using equations 20, differ from the experimentally measured values by less than ±2 per cent.

In test F, as mentioned before, the line has an outer conductor made of a piece of 3-centimeter wave guide while the inner conductor was the same solid rectangular bar used in most of the other tests. The calculated value of the characteristic impedance of this line checks to 1.2 per cent of the measured value. This appears to justify our statement that a very good electrical contact between the four members of the outer conductor of a built-up rectangular line is not very necessary.

In test E, the line is a square coaxial one and as shown in Table II, its characteristic impedance agrees with the measured value to within less than 1 per cent. This proves that the semiempirical formula obtained is valid for any rectangular coaxial line independent of the dimensions of the inside conductor. Moreover, the characteristic impedance was measured as a function of (A/a) and (d/a), that is, as ratios of the outer to the inner conductor dimensions independent of any one dimension.

In applying equations 20, it should be remembered that they have not been checked experimentally for the complete range of the dimensional ratios. On the other hand, however, there is no apparent tendency for the experimental values to deviate from the values calculated by the

Table II. Comparison Between Measured and Calculated Values of the Characteristic Impedance of Rectangular Coaxial Lines

Test Number	Section Number	A	B	a	b	d	A′	d/a	A/a	A/A′	Z_0*	Z_0**	$\frac{\pm\Delta Z_0}{Z_0}\times 100$ per cent
	1	0.446	0.375	0.250	0.0625	0.242	0.446	0.968	1.784	1.000	56.3	56.7	−0.71
	2	0.625	0.375	0.250	0.0625	0.242	0.446	0.968	2.500	1.400	68.0	68.1	−0.15
A	3	0.750	0.375	0.250	0.0625	0.242	0.446	0.968	3.000	1.680	74.4	75.0	−0.80
	4	0.870	0.375	0.250	0.0625	0.242	0.446	0.968	3.500	1.960	79.9	80.0	−0.13
	5	1.000	0.375	0.250	0.0625	0.242	0.446	0.968	4.000	2.240	84.5	84.6	−0.12
	1	0.555	0.500	0.250	0.0625	0.242	0.555	0.968	2.220	1.000	70.8	70.5	+0.40
	2	0.625	0.500	0.250	0.0625	0.242	0.555	0.968	2.500	1.127	74.9	74.2	+0.95
B	3	0.750	0.500	0.250	0.0625	0.242	0.555	0.968	3.000	1.350	81.3	80.4	+1.07
	4	0.875	0.500	0.250	0.0625	0.242	0.555	0.968	3.500	1.577	86.7	87.0	−0.35
	5	1.000	0.500	0.250	0.0625	0.242	0.555	0.968	4.000	1.800	91.4	91.1	+0.33
	1	0.670	0.625	0.250	0.0625	0.242	0.670	0.968	2.680	1.000	82.8	83.5	−0.84
C	2	0.750	0.625	0.250	0.0625	0.242	0.670	0.968	3.000	1.120	86.8	87.1	−0.35
	3	0.875	0.625	0.250	0.0625	0.242	0.670	0.968	3.500	1.307	92.2	93.0	−0.86
	4	1.000	0.625	0.250	0.0625	0.242	0.670	0.968	4.000	1.493	96.8	97.7	−0.92
	1	0.788	0.750	0.250	0.0625	0.242	0.788	0.968	3.152	1.000	92.70	92.00	+0.76
D	2	0.875	0.750	0.250	0.0625	0.242	0.788	0.968	3.500	1.110	96.40	96.30	+0.11
	3	1.000	0.750	0.250	0.0625	0.242	0.788	0.968	4.000	1.270	101.1	102.5	−1.37
	4	1.125	0.750	0.250	0.0625	0.242	0.788	0.968	4.500	1.430	105.2	106.4	−1.13
F	1	0.450	0.200	0.250	0.0625	0.242	0.314	0.968	1.800	1.433	41.50	42.00	−1.20
E	1		0.625	0.250	0.2500	0.000	0.625	0.000	2.500	1.000	53.30	53.10	+0.38

Z_0* = Calculated characteristic impedance (ohms).
Z_0** = Measured characteristic impedance (ohms).
Dimensions are in inches.

formula at the extremes of the range of the ratios studied in this work.

Appendix I. Nomenclature

A = radius of the outer conductor of a circular coaxial line
A = half width of the outer conductor of a rectangular coaxial line
A = semimajor axis of the outer ellipse of an elliptical coaxial line
a = radius of the inner conductor of a circular coaxial line
a = half width of the inside conductor of a rectangular coaxial line
a = semimajor axis of the inside ellipse of an elliptical coaxial line
B = half depth of the outer conductor of a coaxial rectangular line
B = semiminor axis of the outer ellipse of an elliptical coaxial line
b = half depth of the inner conductor of a coaxial rectangular line
b = semiminor axis of the inside ellipse of an elliptical coaxial line
C = shunt capacitance per unit length of a transmission line (farads)
d = distance between the center of an ellipse and either focii
E_{max} = maximum voltage
E_{min} = minimum voltage
f = frequency in cycles or megacycles per second
h = thickness of a resistive coating
L = series inductance per unit length of line (henrys)
R_l = load resistance
R_{eff} = effective resistance
R_{dc} = D-c resistance
V = velocity of propagation of a wave along a transmission line
$VSWR$ = voltage standing wave ratio
Z_l = load impedance (ohms)
Z_0 = characteristic impedance of a transmission line (ohms)
θ = phase angle of the reflection co-efficient
θ_l = load impedance phase angle
λ = wave length
ρ = voltage standing wave ratio
σ = conductivity of a coating material

References

1. Characteristic Impedance of Parallel Wires in Rectangular Troughs, **Sidney Frankel.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 30, number 4, April 1942, pages 182–90.

2. Formulas for the Inductance of Coaxial Busses Composed of Square Tubular Conductors, **H. P. Messinger, T. J. Higgins.** *Electrical Engineering* (*AIEE Transactions*), volume 65, June 1946, pages 328–36.

3. Transmission Lines, Antennas and Wave Guides (book), **King, Minno, Wing.** McGraw-Hill Book Company, New York, N. Y., 1945.

4. Radio Engineering (book), **F. E. Terman.** McGraw-Hill Book Company, New York, N. Y., 1947, page 75.

5. High Frequency Transmission Lines (book), **Willis Jackson.** Methuen and Company, Limited, London, England, 1943, pages 43–46 and page 70.

6. Introduction to Applied Mathematics (book), **F. D. Murnaghan.** John Wiley and Son, Company, New York, N. Y., 1948.

7. Standing Wave Meter, **Heinz E. Kallman.** *Electronics* (New York, N. Y.), volume 20, number 1, January 1947, page 96.

8. Battery Performance from the R.A.C. Power Supply, **George Grammer.** *Q.S.T.*, American Radio Relay League (West Hartford, Conn.), August 1937, page 17.

9. Hyper and Ultra-High Frequency Engineering (book), **R. I. Sarbacher, W. A. Edson.** John Wiley and Son, Company, New York, N. Y., 1943, section 9.21, page 358.

10. The Solution of Transmission-Line Problems by the Use of the Circle Diagram of Impedance, **W. Jackson, L. G. Huxley.** *Journal*, Institution of Electrical Engineers (London, England), volume 91, part III, page 114.

11. Practical Analysis of Ultra-High Frequency Transmission Lines, **J. R. Meagher, H. J. Markly.** RCA Service Company, (Harrison, N. J.) August 1943.

No Discussion

The Calculation of the Magnetizing Force

A. A. HALACSY
MEMBER AIEE

THE author has already published the principles of his new method[1] which is here developed further. This paper presents the developed method in full for the first time.

The method enables one to calculate graphically the magnetizing force H by measuring of areas enclosed by curves derived from the projections of the relevant electric circuit. It also enables one to solve easily such intricate problems as the working out of the values of short circuit forces, or the induction in the case of a complex arrangement of conductors.

The Theory and the Deduction of the Method

The magnetizing force is $\boldsymbol{dH}$ at the point O, see Figure 1, and due to the current I flowing in an element ds of its circuit

$$\boldsymbol{dH}=I(\boldsymbol{ds}\times\boldsymbol{r})/r^3 \qquad (1)$$

This is Biot-Savart's law in vectorial representation. This equation can be transformed, see Appendix I, into

$$\boldsymbol{dH}=2I\boldsymbol{n}dA \qquad (2)$$

where $\boldsymbol{n}$ is the unity vector in the direction of $\boldsymbol{dH}$, that is, normal to the plane of $\boldsymbol{ds}$ and $\boldsymbol{r}$; dA is the area swept over by the radius vector within an angle $d\phi$, see Figure 1. ρ points in the direction of $\boldsymbol{r}$ and has a numerical value $\rho=1/\sqrt{r}$.

$\boldsymbol{dH}$ and $\boldsymbol{ds}$ are infinitesimal quantities. In practical calculation the finite vector of the magnetizing force $\boldsymbol{H}$ is wanted. This $\boldsymbol{H}$ is due to the current I flowing along a finite section s of a circuit, and it can be worked out from $\boldsymbol{dH}$ by using the well-known tools of mathematics as follows

$$\boldsymbol{H}=2I\lim_{dA\to 0}\uparrow\sum_{0}^{s}\boldsymbol{n}dA=2I\uparrow\int_{0}^{s}\boldsymbol{n}dA \qquad (3)$$

The sign ↑ before the sign of the integral calculus should mark and stress the vectoriality of the integration. This must be emphasized because each $\boldsymbol{n}$ may have a different direction. The limits of the integration are set by the beginning and end of the section s.

Equation 3 represents an interesting theoretical result but it is not suitable for practical purposes; a frequent and known drawback of the vectorial representation. The well-known remedy is the introduction of the Cartesian co-ordinates. For this purpose $\boldsymbol{n}$ shall be expressed as the sum of its constituent vector components. (Those are the ones relative to a linearly independent set of the base vectors of the co-ordinate system used.)

$$\boldsymbol{n}=\sum_{h=i}^{m}\boldsymbol{h}\cos\alpha_h \qquad (4)$$

In equation 4, $\boldsymbol{h}$ is the unity vector in the direction h and α_h is the angle between the directions of $\boldsymbol{h}$ and $\boldsymbol{n}$. The combination of equations 3 and 4 results in

$$\boldsymbol{H}=2I\sum_{h=i}^{m}\boldsymbol{h}\int_{0}^{s}dA\cos\alpha_h$$

$$=2I\sum_{h=i}^{m}\boldsymbol{h}\int_{0}^{s}dA_h=2I\sum_{h=i}^{m}\boldsymbol{A}_h \qquad (5)$$

Here $dA_h=dA\cos\alpha_h$ is the projection of the area dA onto a plane normal to the direction $\boldsymbol{h}$. The vector $\boldsymbol{A}_h$ has the numerical value of $\int_{0}^{s}dA_h$ and points in the direction $\boldsymbol{h}$. Equation 5 relieves us of the rather troublesome integration of the divergent vectors $\boldsymbol{n}$ contained in equation 3. Equation 5 enables the calculation by the summation of a finite number of vectors. Furthermore the numerical value of those latter vectors depends on the co-ordinate system we choose. For example, any vector can be expressed by no more than three constituent components pointing into the direction of three ortholinear Cartesian co-ordinates in a 3-dimensional space. Equations 4 and 5 read in this case as follows, see Figure 2

$$\boldsymbol{n}=\boldsymbol{i}\cos\alpha_x+\boldsymbol{j}\cos\alpha_y+\boldsymbol{k}\cos\alpha_z \qquad (6)$$

$$\boldsymbol{H}=2I\left(\boldsymbol{i}\int_{0}^{s}dA_{yz}+\boldsymbol{j}\int_{0}^{s}dA_{zx}+\boldsymbol{k}\int_{0}^{s}dA_{xy}\right) \qquad (7)$$

Equation 7 enables the calculation of $\boldsymbol{H}$ to be made in a very simple way if the areas represented by the scalar integrals could be calculated by some simple means. Let us investigate these areas. dA appears first in the equation 2 and is the area swept over by the radius vector ρ within an angle $d\phi$. Consequently any projection dA_h of that area dA onto a plane normal to a direction $\boldsymbol{h}$, see Figure 3, is swept over by the projection ρ_g of ρ onto the same plane and within the relevant angle $d\phi_g$.

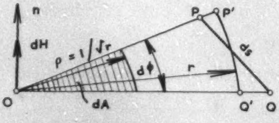


Figure 1. Biot-Savart's law in vectorial representation

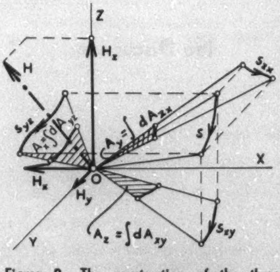


Figure 2. The construction of the three constituent components of H

Paper 52-84, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted February 13, 1951; made available for printing December 27, 1951.

A. A. HALACSY is with the Colonial Development Corporation, Kingstown, St. Vincent, British West Indies.

Table I

No. of point in Figure 5	r_{xy} m	r_{xy}^2 m^2	$r_{xy}^2+d^2$ m^2	$r^{3/2}= (r_{xy}^2+d^2)^{3/4}$ m$^{3/2}$	$\rho_{xy}= r_{xy}/r^{3/2}$ m$^{-1/2}$	r_{yz} m	$\rho_{yz}= r_{yz}/r^{3/2}$ m$^{-1/2}$
0	0	0	0.0064	0.0226	0	0.0800	3.54
1	0.020	0.0004	0.0068	0.0237	0.845	0.0830	3.51
2	0.042	0.00177	0.00817	0.0270	1.55	0.0900	3.34
3	0.064	0.0041	0.0105	0.0330	1.91	0.0985	2.99
4	0.0835	0.0070	0.0134	0.0395	2.12	0.1045	2.64
5	0.099	0.0098	0.0162	0.0455	2.18	0.1070	2.36
6	0.112	0.0126	0.0190	0.0508	2.21	0.1045	2.06
7	0.124	0.0154	0.0218	0.0565	2.20	0.0985	1.74
8	0.133	0.01775	0.02415	0.0618	2.15	0.0900	1.46
9	0.138	0.01908	0.02548	0.0640	2.16	0.0830	1.295
10	0.140	0.01960	0.0260	0.0645	2.17	0.0800	1.240

$d=0.08$ m
$d^2=0.0064$ m^2

$$dA_h=\frac{1}{2}\rho_g{}^2 d\phi_g \tag{8}$$

Naturally all such directions ***g*** are normal to ***h*** because they are on a plane normal to ***h***.

A short calculation, see Appendix II, proves that

$$\rho_g=r_g/r^{3/2} \tag{9}$$

From the combination of equations 5, 8, and 9 we have $\bar{\boldsymbol{H}}$ in its final form, see Appendix III. A component $\boldsymbol{H}_h$ of ***H*** points in the direction ***h*** of the 3-dimensional space and it is

$$\boldsymbol{H}_h=2I\boldsymbol{h}\int_0^{\phi_g}\frac{1}{2}\left(\frac{r_g}{r^{3/2}}\right)^2 d\phi_g \tag{10}$$

Method for the Practical Calculation

Equation 10 presents a simple method of finding the numerical value of a component $\boldsymbol{H}_h$ of the magnetizing force ***H*** pointing from the point *O* into the direction ***h*** and due to the electric current *I* flowing in the section *s* of a circuit. The method is as follows, see Figure 3.

1. Project the section *s* onto the plane passing through the point *O* normal to the direction ***h***.

2. Take all the radii vectors $\boldsymbol{r}_g$ pointing from the point *O* to the points of that s_g projection. Build simultaneously their mates ρ_g collinear with $\boldsymbol{r}_g$ but $\rho_g=r_g/r^{3/2}$ long.

3. Measure the area A_h swept over by ρ_g while ρ_g follows the rotation of r_g from one end point to the other one of the projection s_g.

The requested component of the magnetizing force in the direction ***h*** has the numerical value of

$$H_h=2IA_h \tag{11}$$

All three components of ***H*** pointing into the three directions of the space can be worked out similarly. The total magnetizing force is the vectorial sum of those components, Figure 2. The numerical calculation is very simple if the meter-kilogram-second (mks) system is used.

Advantages of the New Method

The method has the well-known advantages of the graphical calculation. It renders an easy and continuous check of eventual errors as the calculation proceeds. It enables one to predict the effect of an intended alteration in the arrangement of the conductors. It discloses immediately the contribution of any section of a circuit to the resulting magnetizing force. It enables one to calculate the value of a component of the magnetizing force in a deliberately chosen direction, et cetera.

Typical Example

Figures 4 and 5 demonstrate the calculation in the case of a circular loop of $R=0.07$ *m* radius and carrying the current *I*. This case can be handled by simpler methods as well. However, it was chosen for the purpose of demonstration because it enables one to grasp the operational technique of the new method.

H is required at the point *O*, at a distance $d=0.08$ *m* from the loop. The method given should be followed and accordingly three constituent components of ***H*** should be worked out in the *x*, *y*, and *z* direction of 3-dimensional space. First the circuit has to be projected onto the three co-ordinate planes. The projection onto the plane (*xy*) results in a circle with the radius of 0.07 *m*, that is, identical to the circle of the original circuit. The projection onto the planes (*yz*) and (*zx*) respectively result in straight lines. One of those is $AB=0.14$ *m* long and the other one is $CD=0.14$ *m* long.

Secondly we have to measure the distances from the point *O* to the points *P* of the circuit. Figure 4 shows

$$r=\sqrt{(OQ_{xy}+PQ_{xy}{}^2)}=\sqrt{(r_{xy}{}^2+d^2)}$$

where $d=0.08$ *m*, and r_{xy} can be measured from the drawing. On the plane (*xy*) the

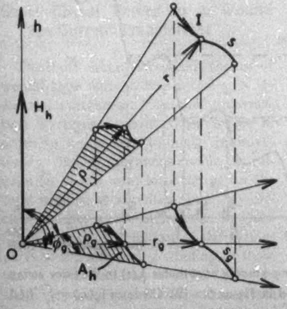


Figure 3. The construction of a component of H in the direction h. The plane of g and s_g is normal to h

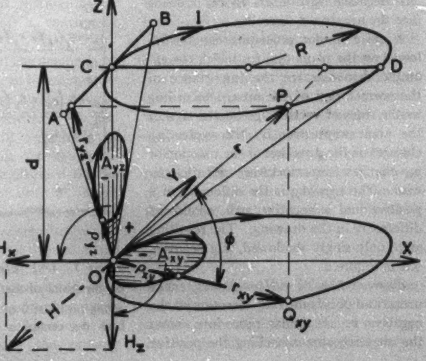


Figure 4 (right). The magnetic field of the current I flowing in a circular loop

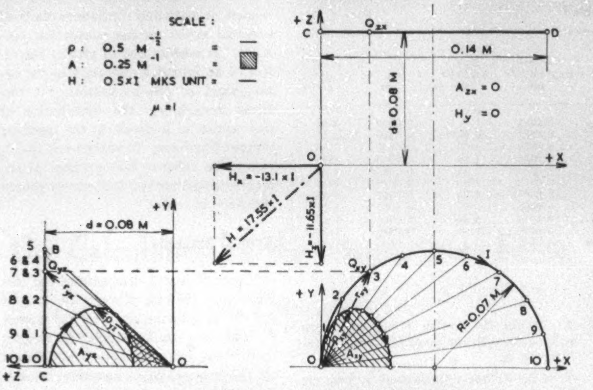


Figure 5. The graphical calculation of H corresponding to Figure 4

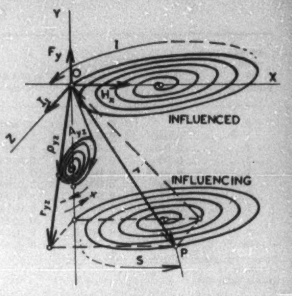


Figure 6. The graphical calculation of the short circuit forces acting on the primary of a current transformer and due to the current of a neighboring current transformer

corresponding values of $\rho_{zy}=r_{zy}/r^2$ are calculated by a slide rule and tabulated in Table I.

The length of each ρ_{xy} calculated in Table I, pointing from O into the corresponding $\mathbf{r}_{xy}$ directions, has to be marked. The ends of these lengths of ρ_{xy} determine the curve circumscribing the area A_{xy}. This area can be measured either by a planimeter or by any other method, for example by counting the covered squares of a graph paper if such one was used for the drawing. The dimension of ρ is $m^{-1/2}$ and that of A is m^{-1}, using the mks system. The scale is given in Figure 5.

According to the rules of vector algebra the sign of the rotation of ρ_{xy} is negative and results in a negative A_{xy}. Due to the symmetrical arrangement it is not necessary to work out more than one-half the complete figure. Thus $A_{xy} = -5.8075$ m^{-1}; $H_z=2IA_{xy}=-11.65I$, mks units. The negative sign means an $\mathbf{H}_z$ pointing into the negative $\mathbf{z}$ direction.

A quite similar procedure can be followed on the plane (yz) for which the example demonstrates the importance of the correct sign of the area. According to the rules of vector algebra the sign of the area swept over by the vector ρ_{yz} changes as the direction of the rotation of ρ_{yz} changes from clockwise to anticlockwise. The area A_{yz} is the difference of a positive and a negative area, shadowed differently in the drawing. The resulting area, only singly shadowed, is a negative one in Figure 5.

$A_{yz}=-6.55m^{-1}$; $H_x=-13.1I$, mks units, and points into the direction of the negative $\mathbf{x}$. A similar procedure shows the negative area cancelling the positive one on the plane (zx). The result is $A_{zx}=0$; $H_y=0$, that is, no component of $\mathbf{H}$ exist in the direction $\mathbf{y}$. This is in agreement with the symmetrical arrangement.

$\mathbf{H}$ is composed of its three components and is $H=\sqrt{(H_x^2+H_y^2+H_z^2)}=I\sqrt{(11.65^2+13.1^2)}=17.55I$, mks units. Its direction can be calculated either by a slide rule or graphically, from the drawing. The calculation took a time of 80 minutes as demonstrated, for 20 points along one-half the circle.

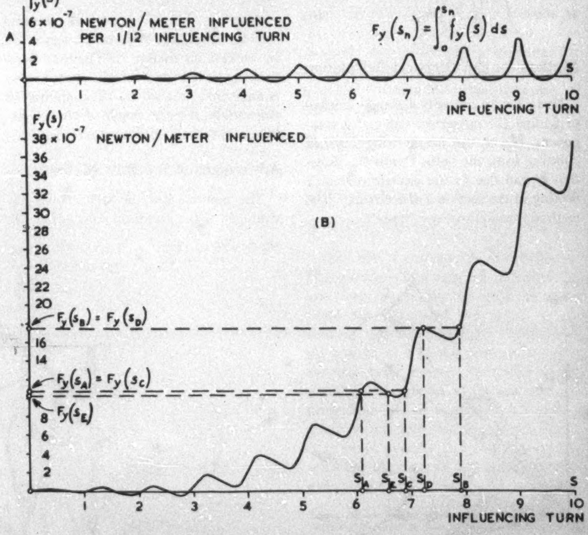


Figure 7. (A) Each 1/12 turn of the influencing winding contributes $f_y(s)$ to the force acting on the point of the influenced winding as marked in Figure 6. (B) The force $F_y(s_n)=\int_0^{s_n} f_y(s)$ acting on point 0 of the influenced current transformer and due to the primary turns s_n (starting from the center) of the influencing current transformer. One ampere is supposed in each winding

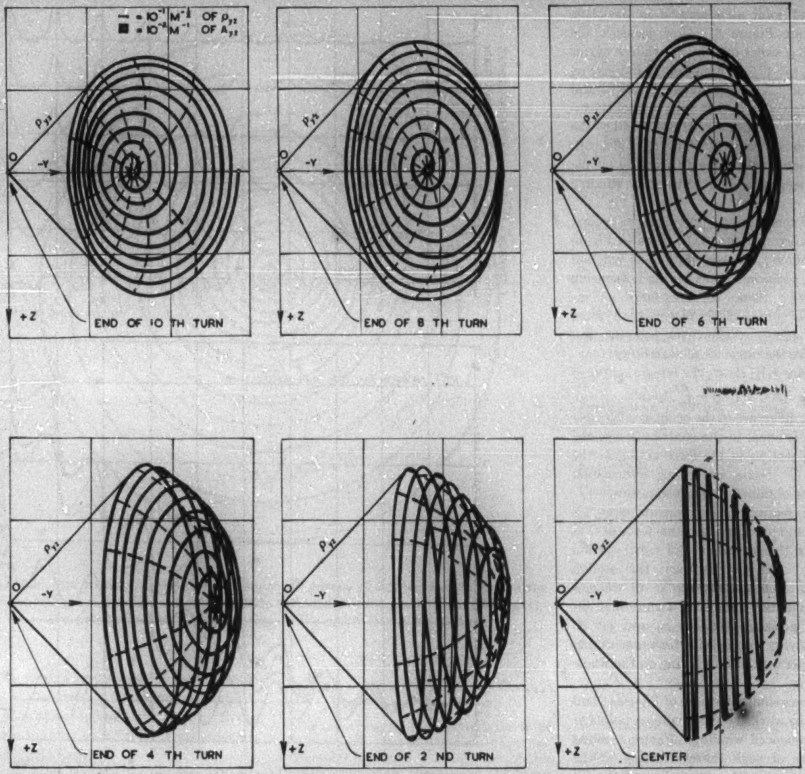


Figure 8 (A). The curves marked by the end point of ρ_{yz} based on the end point of the 1...10 turns of the influenced winding. The end points of the turns are along the axis x in Figure 6. The resulting area of each curve is an A_{yz} related to the point 0 on which ρ_{yz} is based

Short Circuit Forces in a Wound Type Current Transformer

The short circuit forces limit the use of wound type current transformers in industrial switchgear. Detailed information is required about those forces in order to know how to make improvements. Existing formulas enable the calculation of the total force acting on the total winding as an entity, but they disclose neither the distribution of those forces nor the contribution of the various parts of the influencing winding to those forces. Furthermore a small error in those very complex calculations may have a very big influence on the final result and without any possibility of a check. The method presented gives all the required information in a visual representation, with the possibility of a continuous check.

The numerical value of the magnetizing force for any amperage can be reached by multiplying the amperage by the value of the magnetizing force due to 1 ampere. Similarly the numerical value of the mechanical forces equals the value calculated for 1 ampere multiplied by the square of the amperage. For that reason 1 ampere only will be considered here. The presence of the iron cores can be neglected because they become oversaturated in the case of a short circuit.

The value of one component, H_x only, of the magnetizing force is worked out in the following example. This is done for consecutive points along the primary winding of one of two similar current transformers situated in two adjacent phases 0.13 *m* apart in a switchgear, see Figure 6. Both windings are supposed to start from their centers in order to elucidate the situation better, and are wound with a distance of 0.015 *m* between the center lines of the consecutive turns.

The influenced winding carries the current I. The mechanical force F_y acting on a length l of that winding at O and in the field of H_x has a value in Newtons of

$$F_y = 10^{-7} I_1 H_x l \qquad (12)$$

H_x is calculated as in the preceding typical example. Choose O, for example, at the outer end of the influenced winding. ρ_{yz} may be calculated as explained above.

Its end point plots a curve in the plane (yz), see Figure 6. Now measure the area A_{yz} swept over by ρ_{yz} while $\boldsymbol{r}$ moves along, say 1/12 of a turn of the influencing winding. The corresponding mechanical force $f_y(s)$ due to that 1/12 turn of the influencing winding can be calculated by the equations 11 and 12. This $f_y(s)$ is a contribution to the force $f_y(l)$ acting on the unit length of the influenced winding at O.

The consecutive $f_y(s)$ values set against their corresponding 1/12 turns in Figure 7(A) show $f_y(s)$ as a continuous function of the position s taken on the influencing winding. Some sections have an influence considerably bigger than some other ones. Some even oppose and equalize the influence of each other.

Figure 7(B) shows the integral curve of Figure 7(A); $F_y(s_n)=\int_0^{s_n} f_y(s)ds$.

The difference of the ordinates $F_y(s_B)-F_y(s_A)$ is equal to the contribution of the s_B-s_A section to the force acting at the point O. Some interesting information may be obtained from those curves.

The section s_B-s_A spreads over 1.8 turns and produces the same force at O as the six times smaller (0.3 turn) section s_D-s_C. On the contrary the section s_B-s_D produces no force at all whereas the same turn-number in a somewhat different position (between s_D and s_E) effects a very considerable force amounting to 17 per cent of that acting on the whole winding.

Curves similar to those of Figures 6 and 7 can be worked out for various points of the influenced winding. Figure 8 shows a series of such curves relating to the points chosen along a radius of the influenced winding (along the axis x of the Figure 6).

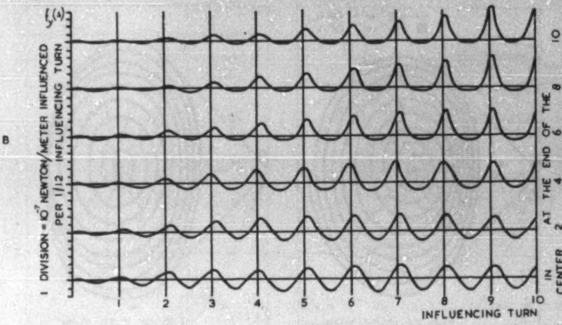
The last ordinate F_y $(s_{\max})$ of each curve, at the right hand end of the figure, represents the total force acting at the point O on the unit length of the related influenced winding; see the explanation in Figure 8(C). O is at a distance l measured from the center and along the influenced winding.

We may introduce the notation F_y $(s_{\max})=f_y(l)$. Figure 9 shows the values $f_y(l)$ against l the latter measured in meter units. The turn-numbers too are indicated in the figure.

$f_y(l)$ encloses the area

$$F_y(l)=\int_0^l f_y(l)=\int_0^l\int_0^{s_{\max}} f_y(s)$$

This is the total force acting on the whole influenced winding. Its numerical value is 16.8×10^{-6} Newton at 1 ampere, or 3.85 pounds at 1,000 amperes, in the present example.



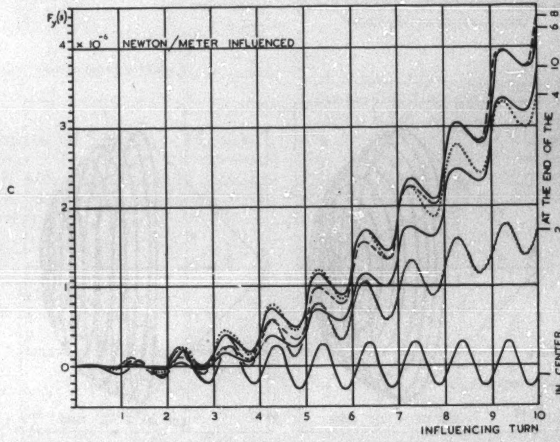
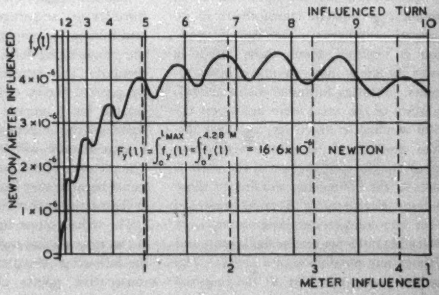


Figure 8 (B). The contribution $F_y(s)$ of 1/12 turns of the influencing primary winding to the force acting on the end point of the 1 . . . 10 turns of the influenced winding

Figure 8 (C). The force $F_y(s_n)$ acting on the end point of the 1 . . . 10 turn of the influenced winding and due to the primary turns s_n of the influencing current transformer. One ampere is supposed in each winding

Figure 9. The force $f_y(l)$ per meter acting distributed on the primary winding of a current transformer and due to the current flowing in a second current transformer similar to the first one and fitted in the neighboring phase. One ampere is supposed in the windings of both current transformers



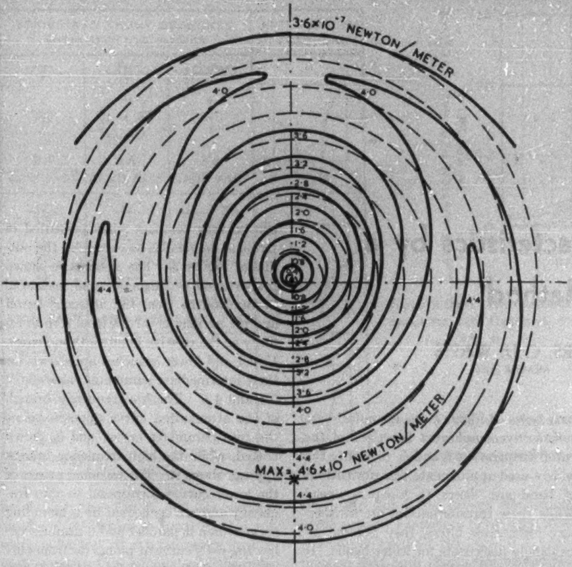


Figure 10. The distribution of the force acting on the primary of a current transformer and due to the current of a similar current transformer fitted in the neighboring phase. The full lines connect points subjected to the same force. The dashed line is the winding

Figure 10 is a map showing the distribution of the forces $f_y(l)$ acting on the unit length's of the influenced winding. Such a map may be useful for the purposes of mechanical design.

A series of such curves and maps shows the influence of the various sections of the influencing winding and gives the most detailed information about the short circuit forces.

Appendix I

The first transformation of the Biot-Savart's law starts from the equation

$$\mathbf{dH}=I(\mathbf{ds}\times\mathbf{r})/r^3 \tag{1}$$

A vector product like $\mathbf{ds}\times\mathbf{r}$ is a vector normal to the plane containing $\mathbf{ds}$ and $\mathbf{r}$. This vector, as any vector, can be written as a product of its unit vector $\frac{\mathbf{ds}}{ds}\times\frac{\mathbf{r}}{r}$ and its numerical value $(ds.r)$

$$\mathbf{ds}\times\mathbf{r}=\left(\frac{\mathbf{ds}}{ds}\times\frac{\mathbf{r}}{r}\right)ds.r \tag{13}$$

The unit vector can be noted by a single letter

$$\mathbf{n}=\frac{\mathbf{ds}}{ds}\times\frac{\mathbf{r}}{r} \tag{14}$$

The numerical value $ds.r$ equals the value of the double area of the elementary triangle OPQ in Figure 1. The area of this triangle approaches the area of the sector $OP'Q'$ if $ds\to O$, which is presumed. Also,

$$ds.r=r^2d\phi \tag{15}$$

Combining equations 1, 13, 14, and 15

$$\mathbf{dH}=I\mathbf{n}d\phi/r \tag{16}$$

Now consider an area dA swept over by a radius vector $\boldsymbol{\rho}$ within an angle $d\phi$ as shown in Figure 1. The vector $\boldsymbol{\rho}$ should point into the direction $\mathbf{r}/r$ and should have the numerical value $\rho=1/\sqrt{r}$, that is

$$\boldsymbol{\rho}=\frac{\mathbf{r}}{r}\frac{1}{\sqrt{r}} \tag{17}$$

The area swept over by this vector $\boldsymbol{\rho}$ is

$$dA=\frac{1}{2}\rho^2d\phi=d\phi/2r \tag{18}$$

Combining equations 16 and 18 we get equation 2

$$\mathbf{dH}=2I\mathbf{n}dA \tag{2}$$

Appendix II

The vector $\boldsymbol{\rho}_g$ is worked out as follows:

Vector algebra shows that the numerical value ρ_g of the projection $\boldsymbol{\rho}_g$ (pointing into the direction $\mathbf{g}$) of the vector $\boldsymbol{\rho}$ equals the scalar product of the original vector $\boldsymbol{\rho}$ and the unit vector $\mathbf{g}/g$, the latter pointing in the direction $\mathbf{g}$.

$$\rho_g=\boldsymbol{\rho}\frac{\mathbf{g}}{g} \tag{19}$$

Combining equations 19 and 17

$$\rho_g=\frac{\mathbf{r}}{r}\frac{1}{\sqrt{r}}\cdot\frac{\mathbf{g}}{g}=\mathbf{r}\cdot\frac{\mathbf{g}}{g}\frac{1}{r^{3/2}} \tag{20}$$

Here it should be noted that $\mathbf{r}\cdot\frac{\mathbf{g}}{g}$ equals the numerical value r_g of the projection $\mathbf{r}_g$ of $\mathbf{r}$ onto the direction $\mathbf{g}$, and consequently

$$\rho_g=r_g/r^{3/2} \tag{9}$$

Furthermore it should be noted that the direction of $\boldsymbol{\rho}_g$ is identical with that of $\mathbf{r}_g$, that is

$$\boldsymbol{\rho}_g/\rho_g=\mathbf{r}_g/r_g \tag{21}$$

Combining equations 21 and 9

$$\boldsymbol{\rho}_g=\mathbf{r}_g/r^{3/2} \tag{22}$$

Appendix III

The combination of equations 5, 8, and 9 results as follows

$$\mathbf{H}=I\sum_{h=i}^{m}\mathbf{h}\int_0^{\phi_g}(r_g^2/r^3)d\phi_g \tag{23}$$

We note that $\mathbf{h}.\mathbf{g}=O$, that is, $\mathbf{g}$ stands in a right angle to $\mathbf{h}$.

In a 3-dimensional space

$$\mathbf{H}=I\left\{\mathbf{i}\int\frac{r_{yz}^2}{r^3}d\phi_i+\mathbf{j}\int\frac{r_{zx}^2}{r^3}d\phi_j+\mathbf{k}\int\frac{r_{xy}^2}{r^3}d\phi_k\right\} \tag{24}$$

$$\mathbf{H}=2I\left\{\mathbf{i}\int\frac{1}{2}\left(\frac{r_{yz}}{r^{3/2}}\right)^2d\phi_i+\mathbf{j}\int\frac{1}{2}\left(\frac{r_{zx}}{r^{3/2}}\right)^2d\phi_j+\mathbf{k}\int\frac{1}{2}\left(\frac{r_{xy}}{r^{3/2}}\right)^2d\phi_k\right\} \tag{25}$$

The component $\mathbf{H}_h$ of $\mathbf{H}$ points in the direction $\mathbf{h}$ and obviously is

$$\mathbf{H}_h=2I\mathbf{h}\int\frac{1}{2}\left(\frac{r_g}{r^{3/2}}\right)^2d\phi_g \tag{10}$$

Reference

1. PRACTICAL CALCULATION OF MAGNETIZING FORCES, A. A. Halacsy. *Proceedings*, Institution of Electrical Engineers (London, England), volume 97, part I, number 104, March 1950, pages 37–42.

No Discussion

Wave Filter Characteristics by a Direct Method

R. C. TAYLOR
ASSOCIATE MEMBER AIEE

MRS. C. U. WATTS
MEMBER AIEE

Synopsis: Filter attenuations for the seven most common ladder filter sections can be read from a single chart or more conveniently from the two charts given. Many properties of low-pass, high-pass, and band-pass filters are readily deduced by inspection. An illustrative example is given and the equations used in construction of the charts are discussed in an Appendix.

THE designer of a filter frequently starts with a pair of cutoff frequencies and the attenuation α in the stop bands on either side of the pass band. To meet these requirements, there is available the basic formula for the propagation constant of a ladder network which is, for a half section:

$$P=\alpha+j\beta=\sinh^{-1}\sqrt{ZY}\,\underline{/\phi} \qquad (1)$$

The design engineer would prefer a simpler expression that would permit P or at least α to be obtained as a function of frequency without the tedious transformations implicit in equation 1. For those who like graphical methods, a chart that is applicable to all ladder networks can be made of the hyperbolic sine as in Figure 1. However, the calculation of $\sqrt{ZY}\,\underline{/\phi}$ with which to enter the chart remains and is a lengthy chore for band-pass filters even though one usually calculates for perfect reactances where ϕ is either zero or 90 degrees. For the constant K low-pass and high-pass sections $\sqrt{ZY}$ is quite easy to calculate but for the various band-pass types of Figure 2 the formulas[1,2] are sufficiently complicated functions of the cutoff frequencies, f_1 and f_2, to justify the widely used approximate transformations of band-pass filters to low-pass ones.[3] While these transformations are satisfactory for narrow filters, they become increasingly inaccurate for wider bands. It is the purpose of this paper to describe a graphical method which yields attenuation without approximation directly as a function of frequency at any relative bandwidth.

Bandwidth as a Parameter

A natural way to describe filter behavior is in terms of bandwidth. For example it is possible to show the attenuation of the band-pass filter section 3A of Figure 2 for all bandwidths as a function of frequency by a family of curves using the parameter, bandwidth divided by geometric mean frequency. Such a chart can be very compact when presented in terms of bandwidth as shown by the outline of Figure 3. The abscissa is shown proportional to frequency and measured in bandwidths from the adjacent cutoff and the parameter p runs from zero to infinity as the relative bandwidth increases. Upper and lower branches are shown for three intermediate parameter values 0.01, 0.1 and 1.0. The low-frequency branch of the attenuation curve appears below the zero parameter curve, and is shown dotted, while the high frequency branch appears above it. The terminal points of the lower curves correspond to zero frequency and are connected by a bounding curve which is parallel and 3 decibels below the $p=0$ curve at points far from cutoff. The abscissa is chosen as the logarithm of bandwidths from cutoff in order to magnify the critical region near cutoff frequency.

Application to Other Sections

Since the performance of the 3B section is identical to that for the 3A, the curves are equally applicable to that section also. As the parameter approaches infinity, both the 3A and 3B sections become 2A sections with cutoff frequency f_2 and the attenuation is found on the $p=\infty$ curve where the bandwidth (f_2-f_1) is f_2 since f_1 becomes zero. Still wider use is suggested

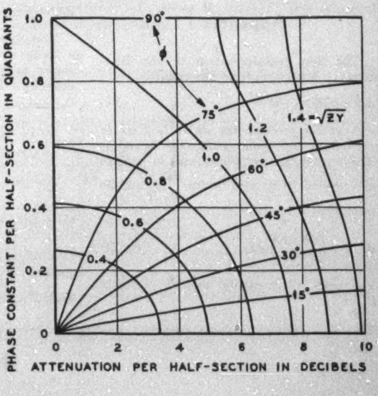
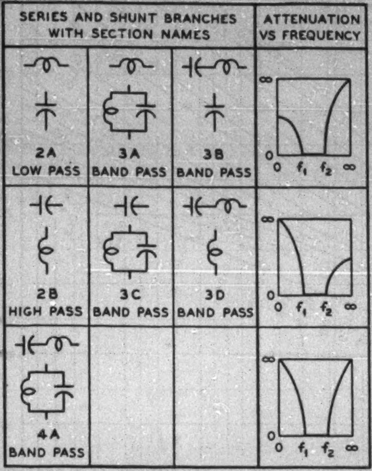


Figure 1. Propagation constant as a function of $\sqrt{ZY}\,\underline{/\phi}$

Paper 52-85, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted October 23, 1951; made available for printing December 17, 1951.

R. C. TAYLOR and MRS. C. U. WATTS are both with the Western Union Telegraph Company, New York, N. Y.

Vo
Pt.
19



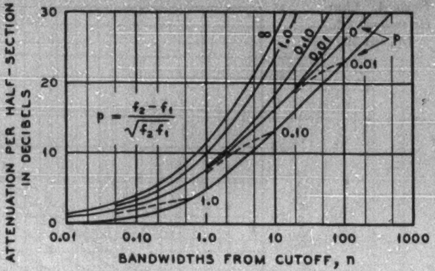


Figure 2 (left). Filter sections associated with Figures 4, 5, and 6

Figure 3 (above). Attenuation as a function of bandwidths from cutoff

by the fact that the 3C and 3D sections have a $\sqrt{ZY}$ which is equal to that for sections 3A and 3B at a frequency inverse with respect to the product of their common cutoff frequencies (equation 2).

$$f_{CD}=f_1f_2/f_{AB} \tag{2}$$

Here the subscript *CD* refers to both section 3C and section 3D of Figure 2, and the subscript *AB* similarly refers to both sections 3A and 3B. This relation is stated more completely in Equation 3.

$$\underset{f=f_{AB}}{(ZY)^{\frac{1}{2}}_{AB}} = \underset{f=f_1f_2/f_{AB}}{(ZY)^{\frac{1}{2}}_{CD}} \tag{3}$$

The 3C and 3D sections with their cutoffs at infinite frequency are 2B high-pass sections with cutoff frequency f_1 and attenuation given by the curve with parameter ∞ using equation 2 which for this limiting case is:

$$f_{2B}=f_1^2/f_{2A} \tag{2A}$$

Since the propagation constant of the 4A section of Figure 2 is the sum of the *AB* and *CD* types, the chart can be read twice and the sum of the readings will give the attenuation constant of the 4A section. In this way, a single chart is sufficient for seven different filter sections. Such a chart is reproduced in detail in Figure 4 with a logarithmic scale extending from 0.01 to 100 bandwidths.

While it is quite true that Figure 4 can be used for this wide variety of cases, the relation of equation 2 results in the same frequency corresponding to two different abscissas. For those who wish to use Figure 4 for all seven cases, Table I is provided; but it has proven worthwhile in the authors' experience to have a second chart to be used for the 3C and 3D sections so that both charts can be entered with the same abscissa which is directly proportional to frequency separation from the adjacent cutoff. This second chart appears as Figure 5. High-pass section 2B requires the inverse equation 2A and Figure 4 since the $p=\infty$ curve cannot be shown in Figure 5.

Phase Constant

Similar charts for β, the imaginary part of the propagation constant, are not necessary in the stop band since for no dissipation, ($\phi=0$ or 90 degrees) β is constant throughout the stop band at a value either 0 or 90 degrees per half section. Inside the pass band where α is zero, β does vary with frequency. Curves of β are plotted against an arithmetic frequency scale in Figure 6 for the limiting cases of the parameter. It is interesting that they are portions of sine waves. It can be shown that if p is as small as 0.02 or as large as 5.0 the phase curves are nearly identical to those for $p=0$ and ∞ respectively. When $p=1$, β is about midway between the limits.

Information Obtainable from the Charts

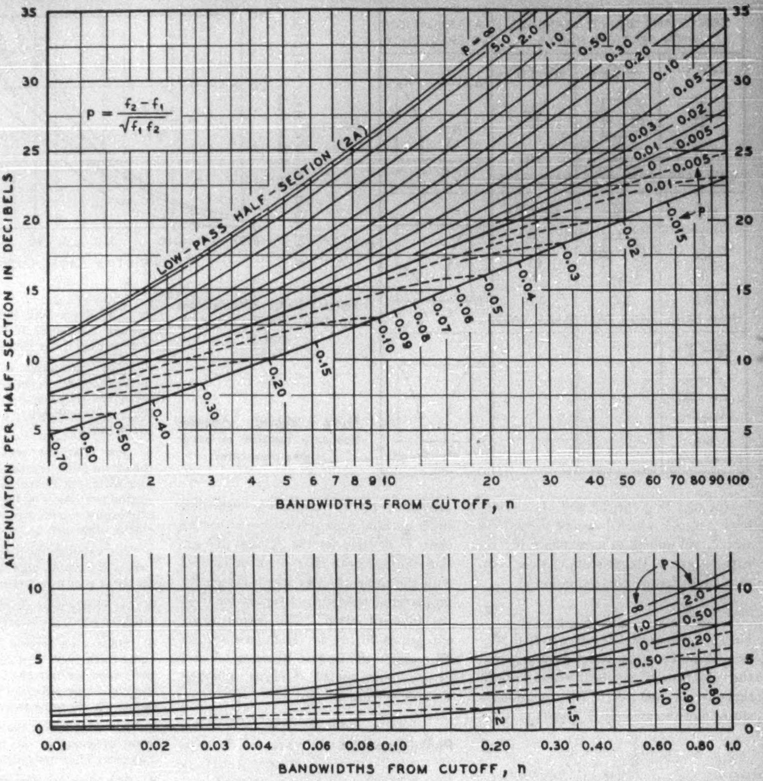
Many properties of wave filters can be easily observed from Figures 3, 4, 5, and 6, for example:

1. The upper and lower branches of 3-element band-pass filters are very nearly symmetrical if the filter is narrow ($p\rightarrow 0$).
2. The attenuation of narrow 3-element filters increases 3 decibels per half section when the distance from cutoff is doubled (that is, 3 decibels per octave or 10 decibels per decade).
3. For relatively wide band-pass filters the attenuation approximates that of a low-pass filter and increases 6 decibels per half section per octave on the side having a steep attenuation curve, and is substantially zero on the other side.
4. For values of the parameter greater than two, a 2-element high or low-pass filter is nearly as good as a 3-element band-pass one.
5. At points near cutoff even rather wide filters ($p=0.5$) are nearly symmetrical.
6. Since a 4A section has an attenuation equal to the sum of a 3AB and a 3CD section, then for narrow bands a 4A section has twice the attenuation of a 3-element section at the same frequency.
7. For wide bands the 4A section has the same attenuation on both sides that a 3-element filter has on its one effective side.
8. The slope of the phase characteristic and therefore the attenuation and time delay in the pass band increase near the cutoff frequencies in a way related to the attenuation in the adjacent stop band.

Other deductions will occur to the user depending on the point of view appropriate to his problem.

M-derived filters and reflection losses are not considered in detail since to do so would duplicate material readily available in the literature. Briefly, Figures 4 and 5 refer to prototype sections only, but a chart or table giving both prototype and derived attenuations as functions of the same quantity can be entered with the prototype attenuation taken from Figures 4 and 5. Reflection losses are a function of matching conditions at the terminals and need not be considered at the intermediate section junctions of a filter where the match is theoretically perfect. At points far from cutoff the terminal reflection losses have 10 decibels per decade slopes.

Figure 4. Chart for sections 2A, 3A, and 3B



Illustrative Example

As an example of the use of this material, Table II is given showing how readily the attenuation of the filter shown schematically in Figure 7 can be obtained. The cutoff frequencies are 600 and 1,000 with no frequency units named since no change in the attenuations of Table II or Figure 7 would result if the cutoffs were 600 and 1,000 cycles per second, kilocycles per second, or for that matter, megacycles per second. The physical form of the filter would, of course, change radically for such large changes in frequency scale. The attenuations of Table II represent about half an hour of work and constitute the basic material needed for any combination of 3A, 3B, 3C, 3D, or 4A band-pass filter sections having a parameter $(f_2-f_1)/\sqrt{f_1 f_2}$ equal to 0.516. The first column consists of enough frequencies to outline the characteristics over the region of interest. They are selected to be commensurate with the bandwidth, in this case 400, thus making the entries in the first bandwidth column round numbers which are used to enter Figures 4 and 5. The 3AB half section attenuations are read from Figure 4, and the 3CD attenuations from Figure 5 or alternatively from Figure 4 using the second column of bandwidths which in general are not round numbers.

The mathematical details underlying Figures 4 and 5 are outlined in the Appendix for the convenience of anyone wishing to construct his own chart.

Table I. Abscissas for Figure 4

Sections	f<f₁	f>f₂
2A, 3A, and 3B	$\frac{f_1-f}{w}$	$\frac{f-f_2}{w}$
3C and 3D	$\frac{f_2}{f}\frac{f_1-f}{w}$	$\frac{f_1}{f}\frac{f-f_2}{w}$
2B	$\frac{f_1}{f}-1$	

Appendix

Attenuation Curves

Equation 1 is the basic formula for obtaining the attenuation of a ladder-type network section. In the stop band when dissipation is neglected, it becomes merely

$$\alpha = \sinh^{-1} \left| \sqrt{ZY} \right|_{\phi=0^\circ} = \cosh^{-1} \left| \sqrt{ZY} \right|_{\phi=90^\circ} \quad (4)$$

The apparent propagation constant, $\sqrt{ZY}$, may be stated in terms of frequency, f, and the two cut-off frequencies, f_1 and f_2, for half sections of types 3A and 3B as follows:

$$(ZY)^{\frac{1}{2}}_{3AB} = \frac{\sqrt{1-\left(\frac{f}{f_1}\right)^2}}{\sqrt{\left(\frac{f_2}{f_1}\right)^2-1}} \quad (5)$$

and for types 3C and 3D by the formula

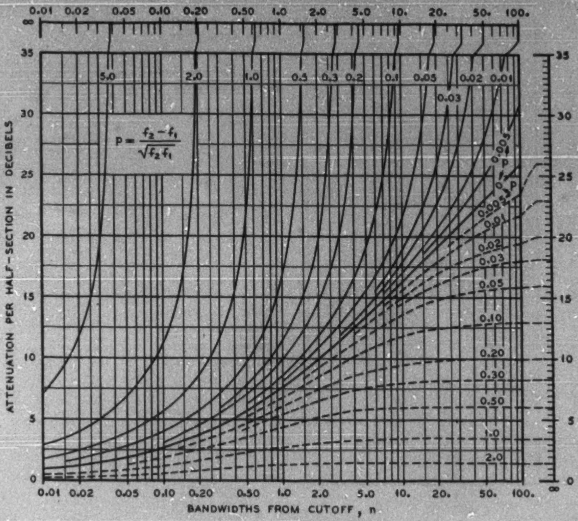


Figure 5. Chart for sections 3C and 3D

$$(ZY)^{\frac{1}{2}}_{3CD} = \frac{\sqrt{1-\left(\frac{f_2}{f}\right)^2}}{\sqrt{\left(\frac{f_2}{f_1}\right)^2-1}} \qquad (5A)$$

The attenuation, α, may also be expressed in terms of bandwidth, $w=f_2-f_1$, and number of bandwidths from cut-off, n. At frequencies above the pass band, $f=f_2+nw$. Therefore

$$\alpha_{3AB} = \cosh^{-1}\left|(ZY)^{\frac{1}{2}}_{3AB}\right| =$$

$$\cosh^{-1}\left[\frac{\sqrt{(1+n)-(1-n^2)\frac{w}{2f_2}}}{\sqrt{1-\frac{w}{2f_2}}}\right] \qquad (6)$$

and

$$\alpha_{3CD} = \sinh^{-1}\left|(ZY)^{\frac{1}{2}}_{3CD}\right| =$$

$$\sinh^{-1}\left[\frac{\left(1-\frac{w}{f_2}\right)\sqrt{n+n^2\frac{w}{2f_2}}}{\left(1+\frac{nw}{f_2}\right)\sqrt{1-\frac{w}{2f_2}}}\right] \qquad (6A)$$

where w/f_2 may be expressed in terms of the parameter, $p=w/\sqrt{f_1f_2}$, as follows:

$$\frac{w}{f_2} = \frac{p^2}{2}\left[\sqrt{1+\left(\frac{2}{p}\right)^2}-1\right] \qquad (7)$$

At frequencies below the pass band, $f=f_1-nw$. Therefore

$$\alpha_{3AB} = \sinh^{-1}\left[\frac{\sqrt{n-n^2\frac{w}{2f_1}}}{\sqrt{1+\frac{w}{2f_1}}}\right] \qquad (8)$$

and

$$\alpha_{3CD} = \cosh^{-1}\left[\frac{1}{\left(1-n\frac{w}{f_1}\right)}\times \frac{\sqrt{(1+n)+(1-n^2)\frac{w}{2f_1}}}{\sqrt{1+\frac{w}{2f_1}}}\right] \qquad (8A)$$

where

$$\frac{w}{f_1} = \frac{p^2}{2}\left[\sqrt{1+\left(\frac{2}{p}\right)^2}+1\right] \qquad (9)$$

Equations 6 and 8 are plotted in decibels in Figure 4, and equations 6A and 8A in Figure 5.

For values of $p>0$, all the curves represented by equation 6 become parallel as n approaches ∞ since

$$\frac{d\alpha}{d(\ln n)} = n\frac{d\alpha}{dn}$$

$$= \frac{\frac{1}{2n}+\frac{w}{2f_2}}{\sqrt{\frac{1}{n}+\frac{w}{2f_2}}\sqrt{\frac{1}{n^2}\left(1-\frac{w}{2f_2}\right)+\frac{1}{n}+\frac{w}{2f_2}}} \qquad (10)$$

and

$$\lim_{n\to\infty} \frac{d\alpha}{d(\ln n)} = 1.0$$

Therefore, for large n's the attenuation at frequencies above the pass band of any 3AB half section increases one neper per power of ϵ—equivalent to 20 decibels per decade.

When $p=0$, that is, for zero bandwidth, formulas 6 and 8A reduce to

$$\alpha = \cosh^{-1}\sqrt{1+n} \qquad (11)$$

while 6A and 8 become

$$\alpha = \sinh^{-1}\sqrt{n} \qquad (12)$$

Since $\cosh^{-1}\sqrt{1+n} = \sinh^{-1}\sqrt{n}$, formulas 11 and 12 are identical. In this case

$$\lim_{n\to\infty} \frac{d\alpha}{d(\ln n)} = 1/2$$

so that for large n's the attenuation increases 10 decibels per decade. This curve separates those to be used above the pass band from those to be used below the pass band in both Figures 4 and 5.

Curves representing attenuations below the pass band must terminate at n's corresponding to zero frequency. This condition is fulfilled when $n=f_1/w$. Substituting n for f_1/w in equation 8,

$$\alpha_{3AB} = \sinh^{-1}\left[\frac{n}{\sqrt{2n+1}}\right] \qquad (13)$$

Equation 13 gives the terminal points for all curves of the family defined by equation 8 and may be considered as the equation of a bounding curve. Each dotted curve of Figure 4 is horizontal where it meets the bounding curve since the derivative of equation 8 is

$$\frac{d\alpha}{d(\ln n)} = 0$$

at $f_1/w=n$.

When n is substituted for f_1/w, that is, at zero frequency, in equation 8A, α_{3CD} becomes infinite. It is obvious that for a given value of the parameter, p, the attenuation curve of Figure 4 meets the bounding curve at the same n for which the corresponding attenuation curve of Figure 5 becomes infinite.

The bounding curve defined by equation 13 lies approximately 3 decibels below the $p=0$ curve defined by equation 12 since

$$\left[\sinh^{-1}\sqrt{n} - \sinh^{-1}\frac{n}{\sqrt{2n+1}}\right] \cong \sinh^{-1}\frac{\sqrt{2}}{4}$$

for large n's. The error introduced by this approximation is only 1 per cent at $n=3$.

Each dotted curve of Figure 5 becomes horizontal as n approaches ∞ since the limit of the derivative of equation 6A is

$$\lim_{n\to\infty} \frac{d\alpha}{d(\ln n)} = 0$$

At this point the attenuation is

$$\alpha_{3CD} = \sinh^{-1}\left[\frac{\frac{f_2}{w}-1}{\sqrt{2\frac{f_2}{w}-1}}\right] \qquad (14)$$

Phase Constant in the Pass Band

In the pass band when zero dissipation is assumed, $\alpha=0$ and $\phi=90$ degrees. Formula 1 then becomes

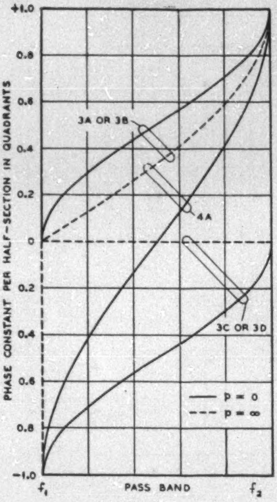


Figure 6 (left). Phase constant as a function of bandwidth for the filter sections of Figure 2

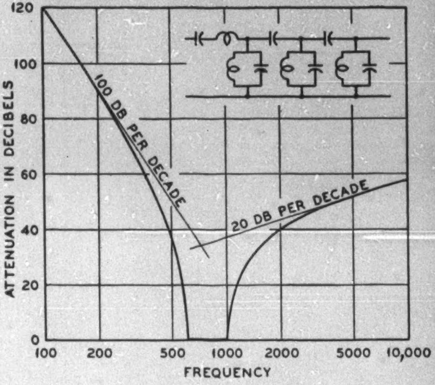


Figure 7 (right). Attenuation and schematic diagram for illustrative band-pass filter

$$\sinh j\beta = j\left|\sqrt{ZY}\right| \tag{15}$$

or

$$\beta = \sin^{-1}\left|\sqrt{ZY}\right| \tag{15A}$$

The relationship between frequency and bandwidth $w=f_2-f_1$ may be written $f=f_1+nw$ where $n<1.0$. Then for 3AB half sections equation 15A becomes

$$\beta = \sin^{-1}\left[\frac{n\sqrt{1+\frac{2f_1}{nw}}}{\sqrt{1+\frac{2f_1}{w}}}\right] \tag{16}$$

so that when $p=0$ (for extremely narrow bands)

$$\beta = \sin^{-1}\sqrt{n} \tag{16A}$$

and when $p=\infty$

$$\beta = \sin^{-1} n \tag{16B}$$

for the low-pass half section 2A.

Similarly for 3CD half sections equation 15A may be written

$$\beta = \sin^{-1}\left[\frac{1}{\left(1+n\frac{w}{f_1}\right)}\frac{\sqrt{2(1-n)-(1-n^2)\frac{w}{f_1}}}{\sqrt{2+\frac{w}{f_1}}}\right] \tag{17}$$

so that when $p=0$ (for extremely narrow bands)

$$\beta = \sin^{-1}\sqrt{1-n} \tag{17A}$$

and when $p=\infty$

$$\beta = 0 \tag{17B}$$

For infinitely wide bands, $p=\infty$, and the curve appears to be on the axes. In this case which is the high-pass half section 2B, an inverse frequency scale in conjunction with the 3AB, $p=\infty$ curve reveals the significant behavior. The limiting β curves for 3A, 3B, 3C, 3D, and 4A half sections are plotted in quadrants in Figure 6.

Table II

Frequency	n-Bandwidths 3AB—Figure 4 3CD—Figure 5	3CD—Figure 4	Decibels Half-Section 3AB a	3CD b	Filter Figure 7 (a+b)+4b
0	1.50	∞	6.0	∞	∞
100	1.25	12.5	5.9	22.7	119.4
200	1.0	5.0	5.7	17.1	91.2
300	0.75	2.5	5.2	13.0	70.2
400	0.5	1.25	4.5	9.8	53.5
520	0.2	0.384	3.1	6.0	33.1
560	0.1	0.178	2.3	4.2	23.3
580	0.05	0.086	1.7	3.0	16.7
*600	0.0	0.0	0.0	0.0	0.0
(Pass Band)					
*1,000	0.0	0.0	0.0	0.0	0.0
1,020	0.05	0.0294	2.2	1.2	8.2
1,040	0.1	0.0577	3.0	1.7	11.5
1,080	0.2	0.111	4.2	2.4	16.2
1,200	0.5	0.25	6.5	3.4	23.5
1,400	1.0	0.428	8.9	4.4	30.9
1,800	2.0	0.667	12.0	5.1	37.5
3,000	5.0	1.0	17.1	5.7	45.6
5,000	10.0	1.20	21.7	5.9	51.2
9,000	20.0	1.33	26.8	6.0	56.8
21,000	50.0	1.43	34.2	6.0	64.2
∞	∞	1.50	∞	6.0	∞

* Cutoff frequencies.

Nomenclature

P = propagation constant per half section
α = attenuation constant in nepers per half section
β = phase constant in radians per half section
Z = series branch impedance of half section
Y = shunt branch admittance of half section
ϕ = phase angle of apparent propagation constant $\sqrt{ZY}$
f = frequency
f_1 = lower cutoff frequency
f_2 = upper cutoff frequency
$w=f_2-f_1$ = bandwidth
$p=w/\sqrt{f_1f_2}$
$n=(f_1-f)/w$ or $(f-f_2)/w$ = bandwidths from cutoff
ϵ = base of natural logarithms
ln = natural logarithm

References

1. Filters for a 150-Kc Carrier System, R. C. Taylor. *AIEE Transactions*, volume 67, part I, 1948, pages 583–88.

2. Transmission Networks and Wave Filters, T. E. Shea. D. Van Nostrand Company, Inc., New York, N. Y., 1929.

3. The Effect of Incidental Dissipation in Filters, E. A. Guillemin. *Electronics* (New York, N. Y.), volume 19, October 1946, pages 130–35.

No Discussion

Series-Connected Magnetic Amplifier With Inductive Loading

THOMAS G. WILSON
ASSOCIATE MEMBER AIEE

THE circuit to be considered in this paper is the series-connected magnetic amplifier with resistance and linear inductance in either the load circuit or the control circuit, or in both circuits simultaneously, Figure 1. This circuit with modifications has been analyzed by many authors, a few of whom are mentioned in references 1 to 9. In almost all the analyses simplifying assumptions have been made concerning the magnitude of some of the circuit parameters, very little work having been done on circuits in which L_a and L_d are not zero. It is the object of this paper to analyze and discuss the changes in current wave shapes produced by variation of the circuit parameters R_a, R_d, L_a, and L_d. In order to understand these changes the circuit is analyzed mathematically. The magnetization constraint of the cores is approximated by three linear segments as shown in Figure 2. The procedure, following a method first used by A. Boyajian,[10] is to obtain solutions for the circuit currents and core fluxes which apply only when the cores are operating on specified linear segments of their respective magnetization curves. When the cores are operating on such a pair of segments the equations of the system become linear, and consequently are solved easily. It is necessary then only to match final and initial conditions at the time of transition from one segment to the next. Equations are thus derived which permit calculation of the instantaneous values of the circuit currents and core fluxes in transient and steady-state conditions. The equations are verified by quantitatively comparing calculated instantaneous values of the current wave shapes with the instantaneous values obtained experimentally. They then are used to explain the changes of wave shape which occur when any of the circuit parameters are changed.

Analysis

The assumptions made in the analysis of the series amplifier of Figure 1 are:

1. The winding resistances of the magnetic amplifier coils and of the external linear inductances may be lumped with the external circuit resistances to determine R_a and R_d.
2. The distributed and interwinding capacitances of the magnetic amplifier coils are negligible.
3. The leakage between the load and control circuit coils is negligible.
4. The constraints linking the flux in a core to the total ampere-turns acting on the core are as given in Figure 2.

Item 4 is particularly applicable for magnetic amplifiers using rectangular magnetization loop core materials, such as thin strip Deltamax and Orthonol, in circuits where the peak currents in the control and load circuits are very large with respect to the magnetizing currents required by the cores. The symmetrical magnetization loop for such a core material is shown by the broken curve in Figure 3.[11] When the circuit is operating as a magnetic amplifier the fluxes in the cores traverse asymmetric magnetization loops such as those shown by the solid curves in Figure 3.[11] The shape of these asymmetric magnetization loops can be obtained only experimentally, and is changed by changes in either the alternating voltage or the direct voltage. It should be noted that even though the asymmetric magnetization loops are mathematically unpredictable all of them lie within a given region, namely, within the large symmetric magnetization loop. Moreover, under usual operating conditions, the peak instantaneous ampere-turns ΣNi on one core exceeds the magnetizing ampere-turns, that is, half the width of the magnetization loop, by several orders of magnitude. For this reason, neglecting the magnetization loop and assuming for the flux versus ampere-turn relation any curve in this region with a positive slope, for example a vertical line, will introduce negligible error in calculations for the currents.

The general voltage equations for the load circuit and the control circuit of Figure 1 are

$$N_a\frac{d\phi_I}{dt}+N_a\frac{d\phi_{II}}{dt}+L_a\frac{di_a}{dt}+R_ai_a=E_a\sin\omega t' \quad (1)$$

and

$$N_d\frac{d\phi_I}{dt}-N_d\frac{d\phi_{II}}{dt}+L_d\frac{di_d}{dt}+R_di_d=E_d \quad (2)$$

These equations are in terms of instantaneous values and must be satisfied at all times. Also the fluxes ϕ_I and ϕ_{II} are constrained at all times to follow the magnetization curves shown in Figure 2. These last two constraints are nonlinear and consequently difficult to handle mathematically when taken in their entirety. However, these magnetization curves are composed of linear segments. Thus, as long as ϕ_I is on a specified segment of its magnetization curve and ϕ_{II} is on a specified segment of its magnetization curve, then the equations of the straight lines through these segments may be used as the magnetization constraints. These two linear equations and the two voltage equations then may be solved by conventional linear theory for the unknown functions of time, i_a, i_d, ϕ_I, and

---

Paper 52-94, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21-25, 1952. Manuscript submitted October 24, 1951; made available for printing December 18, 1951.

THOMAS G. WILSON is with the Naval Research Laboratory, Washington, D. C.

This is a part of the thesis material being submitted in partial fulfillment of the requirements for the degree of Doctor of Science at Harvard University.

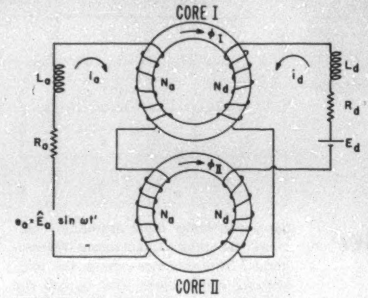


Figure 1 (left). Series-connected magnetic amplifier circuit

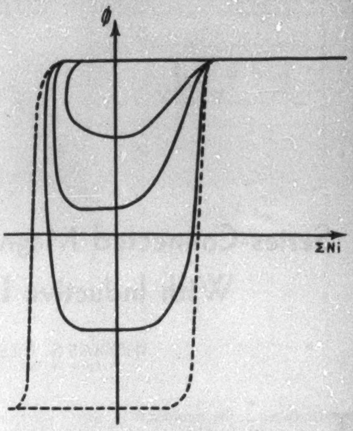


Figure 3 (right). Symmetrical and asymmetrical magnetization loops for a rectangular loop core material

ϕ_{II}. The solutions thus obtained for this mode of operation will apply only as long as neither ϕ_I nor ϕ_{II} exceeds the value corresponding to the end point of its segment. When ϕ_I or ϕ_{II} does reach a value corresponding to the end point of a segment, that is, a knee of its magnetization curve of Figure 2, then it is necessary to use as the magnetization curve constraint the equation of the line through the next segment of the magnetization curve. Operation of the one core on its next adjacent segment and the other core on its original segment constitutes a different mode of operation, and the solutions obtained using this new line as a magnetization curve constraint will apply until either ϕ_I or ϕ_{II} reaches the next knee of its magnetization curve. The complete solutions for the unknowns i_a, i_d, ϕ_I, and ϕ_{II} as functions of time will thus consist of a sequence of solutions, each solution being applicable as long as the fluxes ϕ_I and ϕ_{II} remain on specified segments of their magnetization curves.

Transient

The four possible modes of operation of the series amplifier of Figure 1 are:

Mode I: core I unsaturated and core II unsaturated.

Mode II: core I saturated and core II unsaturated.

Mode III: core I unsaturated and core II saturated.

Mode IV: core I saturated and core II saturated.

These are illustrated in Figure 4. Solutions for the currents and fluxes as functions of time will be obtained for each of these modes of operation.

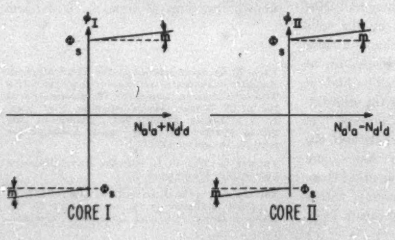
In obtaining the transient response of the amplifier due to step changes in the control voltage, two types of discontinuities occur. One type is a discontinuity due to externally changing the control voltage. The other type is a discontinuity encountered when operation proceeds from one segment of a magnetization curve to the next segment of the same magnetization curve. The mode of operation, however, is changed only at the latter type of discontinuity. Nevertheless, at either type of discontinuity the numerical values in the load and control circuit voltage equations 1 and 2 are different in the interval immediately before and the interval immediately after the discontinuity. As will be seen later, it is more convenient to let θ equal the value of $\omega t'$ at which either type of discontinuity occurs, and to rewrite the load circuit and control circuit voltage equations as

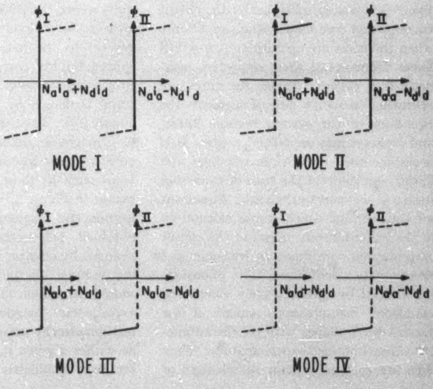
$$N_a\frac{d\phi_I}{dt}+N_a\frac{d\phi_{II}}{dt}+L_a\frac{di_a}{dt}+R_ai_a = E_a\sin(\omega t+\theta) \quad (3)$$

and

Figure 2 (below). Magnetization constraint of the amplifier cores

Figure 4 (right). Operating segments of the magnetization constraint for the four modes of amplifier operation





$$N_a\frac{d\phi_{\text{I}}}{dt}-N_a\frac{d\phi_{\text{II}}}{dt}+L_d\frac{di_d}{dt}+R_di_d=E_d \qquad (4)$$

Writing the a-c supply voltage term as $\mathbf{E}_a\sin(\omega t+\theta)$ instead of $\mathbf{E}_a\sin\omega t'$ serves only to shift the time axis an amount θ/ω seconds whenever either type of discontinuity occurs. Then, in the interval immediately following, t will begin at zero and assume positive values until the next discontinuity occurs. Here the time axis again is shifted a new value θ/ω seconds where θ is now the value of $\omega t'$ at this discontinuity, that is, θ equals the value of θ corresponding to the preceding discontinuity plus ω radians per second multiplied by the time in seconds between the two discontinuities. The values of ϕ_{I}, ϕ_{II}, and E_d corresponding to the next interval are substituted in equations 3 and 4 and the process repeated. Thus, to obtain the solutions for currents and fluxes in any mode of operation, it is necessary only to solve equations 3 and 4 simultaneously with the equations of the lines through the segments of the magnetization curve concerned.

Mode I

In Mode I both cores are unsaturated, that is, both cores are operating on the vertical segments of their magnetization curves. Referring to Figure 2 the equations of these lines are

$$N_ai_a+N_di_d=0 \text{ (core I)} \qquad (5)$$

and

$$N_ai_a-N_di_d=0 \text{ (core II)} \qquad (6)$$

These two constraint equations and the two voltage equations 3 and 4 must all be satisfied simultaneously. The simultaneous solutions of equation 5, 6, 3, and 4 are

$$i_a=0 \qquad (7)$$

$$i_d=0 \qquad (8)$$

$$\phi_{\text{I}}=\frac{\mathbf{E}_a}{\omega 2N_a}[\cos\theta-\cos(\omega t+\theta)]+\frac{E_d}{2N_d}t+\phi_{\text{I}}(0) \qquad (9)$$

and

$$\phi_{\text{II}}=\frac{\mathbf{E}_a}{\omega 2N_a}[\cos\theta-\cos(\omega t+\theta)]-\frac{E_d}{2N_d}t+\phi_{\text{II}}(0) \qquad (10)$$

where $\phi_{\text{I}}(0)$ and $\phi_{\text{II}}(0)$ are, respectively, the values of ϕ_{I} and ϕ_{II} at $t=0$ in equations 9 and 10 or the values of ϕ_{I} and ϕ_{II} at θ in the a-c supply voltage. These solutions will apply whenever both ϕ_{I} and ϕ_{II} have values between $+\Phi_s$ and $-\Phi_s$.

Mode II

In Mode II core I operates along the upper saturated segment of its magnetization curve and core II operates along the unsaturated vertical segment of its magnetization curve. The magnetization curve constraints can be written as the equations of the lines through these segments or

$$\phi_{\text{I}}=m(N_ai_a+N_di_d)+\Phi_s \text{ (core I)} \qquad (11)$$

and

$$N_ai_a-N_di_d=0 \text{ (core II)} \qquad (12)$$

Equations 11 and 12 and the voltage equations 3 and 4 must all be satisfied simultaneously. The solutions are

$$i_a=\frac{\mathbf{E}_a}{\sqrt{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}}\times\left\{\sin(\omega t+\theta-\alpha)-\sin(\theta-\alpha)\times\exp\left[-\frac{(R_a+R_d')t}{L_a'+L_d'}\right]\right\}+\frac{N_a}{N_d}\times\frac{E_d}{R_a+R_d'}\left\{1-\exp\left[-\frac{(R_a+R_d')t}{L_a'+L_d'}\right]\right\}+i_a(0)\exp\left[-\frac{(R_a+R_d')t}{L_a'+L_d'}\right] \qquad (13)$$

$$i_d=\frac{N_a}{N_d}i_a \qquad (14)$$

$$\phi_{\text{I}}=2mN_ai_a+\Phi_s \qquad (15)$$

$$\phi_{\text{II}}=\frac{\mathbf{E}_a}{\omega N_a}\frac{R_d'}{R_a+R_d'}\cos\theta-\frac{\mathbf{E}_a}{\omega N_a}\sqrt{\frac{(R_d')^2+(\omega L_d')^2}{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}}\times\cos(\omega t+\theta-\alpha+\beta)-\frac{1}{N_a}\frac{L_d'R_a-L_a'R_d'}{R_a+R_d'}\times\left\{\frac{\mathbf{E}_a}{\sqrt{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}}\times\sin(\theta-\alpha)\exp\left[-\frac{(R_a+R_d')t}{L_a'+L_d'}\right]-\left[\frac{N_a}{N_d}\frac{E_d}{R_a+R_d'}-i_a(0)\right]\times\left[\!\left[1-\exp\left[-\frac{(R_a+R_d')t}{L_a'+L_d'}\right]\right]\!\right]\right\}-\frac{E_d}{N_d}\frac{R_a}{R_a+R_d'}t+\phi_{\text{II}}(0) \qquad (16)$$

Here $i_a(0)$, $\phi_{\text{I}}(0)$, and $\phi_{\text{II}}(0)$ are the values of i_a, ϕ_{I}, and ϕ_{II} at $t=0$ in the equations or at θ in the a-c supply voltage. The other terms are defined in Appendix II. These solutions will apply as long as ϕ_{I} remains on its upper saturated segment and ϕ_{II} remains simultaneously on its vertical unsaturated segment.

Mode III

In Mode III core I operates on its unsaturated vertical segment and core II operates on its lower saturated segment. The equations of these constraints are

$$N_ai_a+N_di_d=0 \text{ (core I)} \qquad (17)$$

and

$$\phi_{\text{II}}=m(N_ai_a-N_di_d)-\Phi_s \text{ (core II)} \qquad (18)$$

The simultaneous solutions of equations 17, 18, 3, and 4 are the same as equations 13, 14, 15, and 16 if everywhere E_d is replaced by $-E_d$. The solutions so obtained will apply as long as core I remains on its unsaturated vertical segment and core II on its lower saturated segment.

Mode IV

In this mode both cores are saturated, core I operating on its upper saturated segment and core II operating on its lower saturated segment. The equations describing these constraints are

$$\phi_{\text{I}}=m(N_ai_a+N_di_d)+\Phi_s \text{ (core I)} \qquad (19)$$

and

$$\phi_{\text{II}}=m(N_ai_a-N_di_d)-\Phi_s \text{ (core II)} \qquad (20)$$

The general voltage equations 3 and 4 also must be satisfied in this interval. The solutions for the load and control circuit currents are obtained by simultaneously solving equations 19, 20, 3, and 4. They are

$$i_a=\frac{\mathbf{E}_a}{\sqrt{R_a^2+(\omega L_a')^2}}\times\left\{\sin(\omega t+\theta-\gamma)-\sin(\theta-\gamma)\times\exp\left[-\frac{R_at}{L_a'}\right]\right\}+i_a(0)\exp\left[-\frac{R_at}{L_a'}\right] \qquad (21)$$

$$i_d=\frac{E_d}{R_d}\left\{1-\exp\left[-\frac{R_d't}{L_d'}\right]\right\}+i_d(0)\exp\left[-\frac{R_d't}{L_d'}\right] \qquad (22)$$

These solutions show that when the amplifier is operating in Mode IV the instantaneous ampere-turns of the control circuit are no longer required to equal the instantaneous ampere-turns of the load circuit.

Calculation Procedure

The solutions for all four modes of operation have been summarized in Table I. To use this table in calculating the instantaneous values of the transient currents flowing in response to a step change in the control voltage E_d, it is necessary to know only the values of the various currents and fluxes at the instant E_d is changed. As an example, assume that the value of the a-c supply voltage is such that, when acting alone in the circuit, it causes the flux in both cores to swing between $+\Phi_s$ and $-\Phi_s$, or $\mathbf{E}_a=\omega 2N_a\Phi_s$ and suppose that it is desired to find the response of the amplifier when the control voltage E_d is suddenly increased from 0 to

Table I

Mode	Core I	Core II	Solutions
I	U	U	$i_a=0$ $i_d=0$ $\phi_I=\frac{E_a}{\omega 2N_a}[\cos\theta-\cos(\omega t+\theta)]+\frac{E_d}{2N_d}t+\phi_I(0)$ $\phi_{II}=\frac{E_a}{\omega 2N_a}[\cos\theta-\cos(\omega t+\theta)]-\frac{E_d}{2N_d}t+\phi_{II}(0)$
II	S	U	$i_a=\frac{E_a}{\sqrt{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}}\left\{\sin(\omega t+\theta-\alpha)-\sin(\theta-\alpha)\times\exp\left[-\frac{(R_a+R_d')t}{L_a'+L_d'}\right]\right\}+\frac{N_a}{N_d}\frac{E_d}{R_a+R_d'}\left\{1-\exp\left[-\frac{(R_a+R_d')t}{L_a'+L_d'}\right]\right\}+i_a(0)\exp\left[-\frac{(R_a+R_d')t}{L_a'+L_d'}\right]$ $i_d=\frac{N_a}{N_d}i_a$ $\phi_I=2mN_ai_a+\Phi_s$ $\phi_{II}=\frac{E_a}{\omega N_a}\frac{R_d'}{R_a+R_d'}\cos\theta-\frac{E_a}{\omega N_a}\sqrt{\frac{(R_d')^2+(\omega L_d')^2}{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}}\cos(\omega t+\theta-\alpha+\beta)-\frac{1}{N_a}\frac{L_d'R_a-L_a'R_d'}{R_a+R_d'}\left\{\frac{E_a}{\sqrt{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}}\times\sin(\theta-\alpha)\exp\left[-\frac{(R_a+R_d')t}{L_a'+L_d'}\right]-\left[\frac{N_a}{N_d}\frac{E_d}{R_a+R_d'}-i_a(0)\right]\times\left[\!\left[1-\exp\left[-\frac{(R_a+R_d')t}{L_a'+L_d'}\right]\right]\!\right]\right\}-\frac{E_d}{N_d}\frac{R_a}{R_a+R_d'}t+\phi_{II}(0)$
III	U	S	$i_a=\frac{E_a}{\sqrt{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}}\left\{\sin(\omega t+\theta-\alpha)-\sin(\theta-\alpha)\times\exp\left[-\frac{(R_a+R_d')t}{L_a'+L_d'}\right]\right\}-\frac{N_a}{N_d}\frac{E_d}{R_a+R_d'}\left\{1-\exp\left[-\frac{(R_a+R_d')t}{L_a'+L_d'}\right]\right\}+i_a(0)\exp\left[-\frac{(R_a+R_d')t}{L_a'+L_d'}\right]$ $i_d=-\frac{N_a}{N_d}i_a$ $\phi_I=\frac{E_a}{\omega N_a}\frac{R_d'}{R_a+R_d'}\cos\theta-\frac{E_a}{\omega N_a}\sqrt{\frac{(R_d')^2+(\omega L_d')^2}{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}}\times\cos(\omega t+\theta-\alpha+\beta)-\frac{1}{N_a}\frac{L_d'R_a-L_a'R_d'}{R_a+R_d'}\left\{\frac{E_a}{\sqrt{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}}\times\sin(\theta-\alpha)\exp\left[-\frac{(R_a+R_d')t}{L_a'+L_d'}\right]+\left[\frac{N_a}{N_d}\frac{E_d}{R_a+R_d'}+i_a(0)\right]\times\left[\!\left[1-\exp\left[-\frac{(R_a+R_d')t}{L_a'+L_d'}\right]\right]\!\right]\right\}+\frac{E_d}{N_d}\frac{R_a}{R_a+R_d'}t+\phi_I(0)$ $\phi_{II}=2mN_ai_a-\Phi_s$
IV	S	S	$i_a=\frac{E_a}{\sqrt{R_a^2+(\omega L_a')^2}}\left\{\sin(\omega t+\theta-\gamma)-\sin(\theta-\gamma)\exp\left[-\frac{R_at}{L_a'}\right]\right\}+i_a(0)\exp\left[-\frac{R_at}{L_a'}\right]$ $i_d=\frac{E_d}{R_d}\left\{1-\exp\left[-\frac{R_d't}{L_d'}\right]\right\}+i_d(0)\exp\left[-\frac{R_d't}{L_d'}\right]$ $\phi_I=m(N_ai_a+N_di_d)+\Phi_s$ $\phi_{II}=m(N_ai_a-N_di_d)-\Phi_s$

A volts at θ_0 in the a-c supply voltage wave. With zero control voltage the amplifier will operate in Mode I in the steady state. The a-c supply voltage will be divided equally between the two amplifier coils and no currents will flow. The fluxes ϕ_I and ϕ_{II} will be sinusoidal, lagging the a-c supply voltage by 90 degrees, and will swing between $+\Phi_s$ and $-\Phi_s$. This steady-state condition may be expressed by setting $\theta=\pi/2$ radians, $E_d=0$ volts, $\phi_I(0)=0$ webers, and $\phi_{II}(0)=0$ webers in the equations given in Mode I of Table I. These equations then become $i_a=0$, $i_d=0$, $\phi_I=\phi_{II}=-\Phi_s\cos(\omega t+\pi/2)$ and will apply until the control voltage is suddenly increased to A volts at θ_0 in the a-c supply voltage wave. At this instant, $\omega t+\pi/2$ in the above flux equation equals θ_0. In the interval immediately after θ_0 the amplifier will continue operating in Mode I, but the control voltage E_d will now have the value A.

The equations for this interval are given in Mode I of Table I where now $\theta=\theta_0$, $E_d=A$, $\phi_I(0)=-\Phi_s\cos\theta_0$, and $\phi_{II}(0)=-\Phi_s\cos\theta_0$. It can be seen from the equations that if the a-c supply voltage were positive at θ_0 then both ϕ_I and ϕ_{II} will increase. However, the rate of increase of ϕ_I will be greater than the rate of increase of ϕ_{II} by an amount directly proportional to A, and core I will saturate before the a-c supply voltage reaches zero; see Figure 5.

Let θ_1 be the time in the a-c supply voltage wave when ϕ_I reaches the value Φ_s. At this instant the currents and fluxes are given by the equations in Mode I of Table I when $\omega t=\theta_1-\theta_0$. These values will be substituted for $i_a(0)$, $\phi_I(0)$, and $\phi_{II}(0)$ in the equations which apply for the next mode of operation, which is Mode II in Table I. In the equations for this mode $\theta=\theta_1$, $E_d=A$, and it can be seen that the currents i_a and i_d will increase from zero to a maximum and then decrease again to zero. In the meantime the fluxes ϕ_I and ϕ_{II} will increase and then decrease, ϕ_I increasing from Φ_s along its upper saturated segment and decreasing along the same segment to ϕ_s, and ϕ_{II} increasing and decreasing along its vertical unsaturated segment. The condition determining when operation in this mode ceases is that the flux ϕ_I returns to the value Φ_s. Let this time in the a-c supply voltage wave be θ_2. In the interval after θ_2 both cores will be unsaturated. This represents the conditions for Mode I. The a-c supply voltage, however, is now negative and both ϕ_I and ϕ_{II} will decrease, ϕ_{II} decreasing at a rate faster than ϕ_I by an amount proportional to A. The equations for this interval are given in Mode I of Table I where $\theta=\theta_2$, $E_d=A$, $\phi_I(0)=\Phi_s$, and $\phi_{II}(0)$ equals the value of ϕ_{II} at $\omega t=\theta_2-\theta_1$ determined from its Mode II equation. This mode of operation ends when $\phi_{II}=-\Phi_s$ at θ_3 in the a-c supply voltage wave.

The next mode of operation is Mode III in Table I. In this mode i_a decreases from zero to a minimum and then increases to zero, i_d increases from zero to a maximum and then decreases to zero, ϕ_I decreases along its vertical unsaturated segment from its value at θ_3 to a minimum and then increases, and ϕ_{II} decreases from $-\Phi_s$ along its lower saturated segment to a minimum and then increases to $-\Phi_s$. In the equations for this mode $i_a(0)$, $\phi_I(0)$, and $\phi_{II}(0)$ are the values of these parameters at $\omega t=\theta_3-\theta_2$ in the Mode I equations. Operation in this mode ends when ϕ_{II} returns to the value $-\Phi_s$ at θ_4 in the a-c supply voltage wave.

The next mode again will be Mode I since the conditions in the circuit are similar to those which existed in the in-

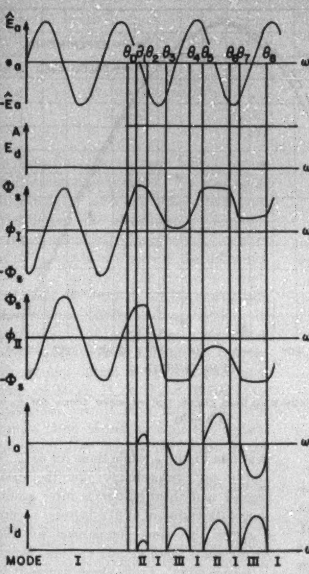

$\hat{E}_a$
e_a
$-\hat{E}_a$
θ_0 θ_1 θ_2 θ_3 θ_4 θ_5 θ_6 θ_7 θ_8
ωt'
A
E_d
ωt'
Φ_s
ϕ_I
$-\Phi_s$
ωt'
Φ_s
ϕ_{II}
$-\Phi_s$
ωt'
i_a
ωt'
i_d
ωt'
MODE I II I III I II I III I

Figure 5 (left). Sketch of transient wave shapes

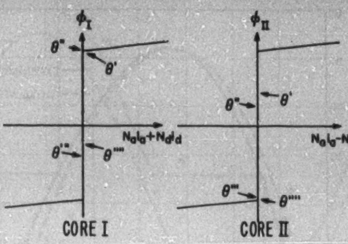


Figure 6 (right). Location of steady-state θ's on the magnetization constraint curves

terval θ_0 to θ_1. The cycle of operation outlined here from θ_0 to θ_4 constitutes the most common sequence of the modes, and over the normal operating range the cycle will repeat until a steady state is reached. Under certain circumstances, however, this cycle may repeat only a few times and then a different cycle, composed of fewer modes or a new mode, may begin.

For example, consider what would happen when the amplifier operates in Mode II after having completed a few of the cycles. In this mode core I operates on its upper saturated segment while core II operates on its vertical unsaturated segment. As stated previously ϕ_{II} will increase to a maximum and then decrease. It is possible for ϕ_{II} to reach the lower knee of its magnetization curve while ϕ_I is still operating on its upper saturated segment. Under this condition the next mode of operation will be Mode IV and a new cycle of operation will begin.

Throughout this section it will be noticed that in order to perform a numerical calculation of the transient response of the amplifier it is not necessary to know what modes of operation will occur or the sequence in which the modes occur. It is only necessary to know the values of the two currents and the two fluxes at the instant the transient is initiated. With these values known the complete transient may be calculated by continuously plotting ϕ_I and ϕ_{II}, since these two functions completely determine the mode of operation.

Steady State

The transient analysis outlined gives a method whereby the instantaneous values of the currents and fluxes may be calculated as functions of time if the values of the currents and fluxes are known at the instant the control voltage is changed. The complete solution was seen to consist of a sequence of solutions which apply during different modes of operation. The order in which these modes of operation occur was determined by continuously plotting the two flux waves ϕ_I and ϕ_{II}. The instantaneous values of the current and flux waves thus could be plotted until a steady state is reached. Thus the equations given in Table I for the various modes will apply also to the steady state since the latter is only a special case of the transient.

In the normal control range the order of the modes of operation of the series amplifier is Mode II, Mode I, Mode III, Mode I, and repeat. Let θ', θ'', θ''', and θ'''' be the angles in the a-c supply voltage wave at which operation in the respective modes begins. The fluxes at these times are indicated in Figure 6 for a steady-state condition. It should be noticed in the equations for Mode II, Table I, that operation in this mode ceases when ϕ_I decreases to a value Φ_s, or when $i_a=0$. Thus if θ' is known, θ'' can be determined by setting $i_a=0$. Or, one relation between θ' and θ'' which is always true is

$$\sin(\theta''-\alpha)=\sin(\theta'-\alpha)\times \exp\left[-\frac{(R_a+R_d')(\theta''-\theta')}{\omega(L_a'+L_d')}\right]- \frac{N_a}{N_d}\frac{E_d}{\mathbf{E}_a}\frac{\sqrt{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}}{R_a+R_d'}\times \left\{1-\exp\left[-\frac{(R_a+R_d')(\theta''-\theta')}{\omega(L_a'+L_d')}\right]\right\} \quad (23)$$

From Figure 6 it can be seen that in the steady state the change in flux in core I from θ'' to θ''' must equal the change in flux in core II from θ' to θ''''. Also θ''' must be equal to $\theta'+\pi$ radians. These facts may be used to obtain the following simple steady-state relation between θ' and θ''.

$$\cos\theta'-\frac{N_a}{N_d}\frac{E_d}{\mathbf{E}_a}\theta'=\cos\theta''-\frac{N_a}{N_d}\frac{E_d}{\mathbf{E}_a}\theta''+ \frac{R_a+R_d'}{R_d'}\frac{N_a}{N_d}\frac{E_d}{\mathbf{E}_a}\pi \quad (24)$$

The derivation of this equation is given in Appendix I. The two transcendental equations 23 and 24 may be solved simultaneously by graphical means to determine the angles θ' and θ''. These two angles completely specify the steady state since $\theta'''=\theta'+\pi$ and $\theta''''=\theta''+\pi$. They may be substituted in the equations of their respective modes to obtain the instantaneous values of the steady-state wave shapes.

Experimental Verification

Before drawing conclusions from this analysis it was deemed necessary to check the validity of the assumptions made. One circuit used to compare analytical and experimental results is shown in Figure 1 wherein

N_a = 1,000 turns
R_a = 109.3 ohms

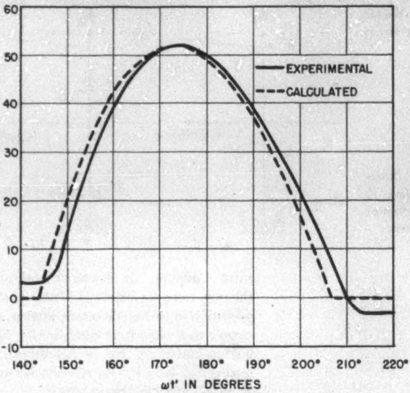


Figure 7. Initial steady-state load circuit current wave for $E_d = 1.32$ volts

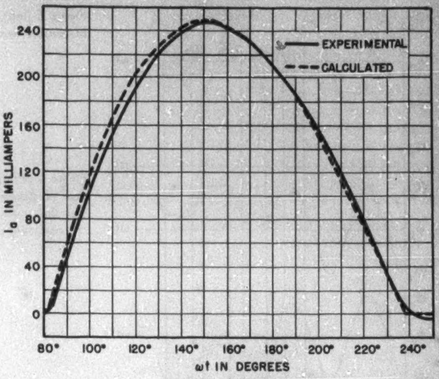


Figure 9. Final steady-state load circuit current wave shape for $E_d = 15.3$ volts

$L_a = 0.221$ henry
$E_a = 79.7$ volts
$N_d = 1{,}000$ turns
$R_d = 110.9$ ohms
$L_d = 0.170$ henry
$\Phi_s = 1.015\times10^{-4}$ weber
$m = 4.9\times10^{-8}$ weber per ampere-turn

Cores are made of Deltamax

The input disturbance used was a step which changed the control voltage E_d from 1.32 volts to 15.3 volts at 212 degrees in an alternating voltage cycle. Figure 8(A) is an oscillogram of this transient showing, from top to bottom, the step control circuit voltage, the sinusoidal load circuit voltage, the load circuit current, and the control circuit current. Figure 7 shows a comparison between the results of a steady-state calculation for the load circuit current and the corresponding experimental current obtained from Figure 8(A). Both are for a control voltage $E_d = 1.32$ volts. Figure 8(B) compares the calculated and experimental transient load circuit current waves for the conditions just stated. In this figure, the waves for $n = 1$ are the same as shown in Figure 7, while those for $n = 2$, 3, and 4 are, respectively, for the first, second, and third half-cycle after initiation of the transient at 212 degrees. Figure 9 shows the same comparison for the steady-state load circuit currents when $E_d = 15.3$ volts. From these figures it is seen that the calculated and experimental current waves are in very close agreement. Since appropriate matching was obtained between the instantaneous values determined analytically and those determined experimentally, it is assumed that the mathematical model closely represents the physical system. With this established, conclusions concerning the operation of the magnetic amplifier now will be drawn from examination of the mathematical model.

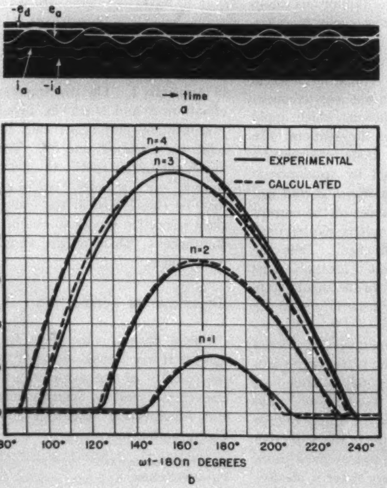


Figure 8 (A). Oscillogram of transient currents and voltages when E_d is increased from 1.32 volts to 15.3 volts. (B). Initial steady-state load circuit current wave when $E_d = 1.32$ volts ($n = 1$) and the first three half-cycles ($n = 2$, 3, 4) of the transient current wave, when E_d is changed from 1.32 volts to 15.3 volts at 212 degrees

Effect of Parameter Variation

This section discusses the changes produced in the steady-state current wave shapes when the circuit parameters are varied. This is done by examining the equations established in the mathematical analysis to see the effect of changing a single parameter while holding constant the remaining parameters.

As stated previously, when the amplifier is operating throughout its normal control range the steady-state angles at which the currents begin flowing and cease flowing are determined by simultaneously solving equations 23 and 24. If these two equations were used in their entirety, it would be very difficult to visualize the effect on θ' and θ'' brought about by changing one of the circuit parameters. To simplify the following discussion it will be assumed that the exponential terms in equation 23 are negligible. Equation 23 thus reduces to

$$\sin(\theta''-\alpha)=-\frac{\sqrt{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}}{R_a+R_d'}\,\frac{N_a}{N_d}\,\frac{E_d}{E_a} \qquad (25)$$

In many instances the error introduced by this assumption is small. By way of illustration the error in θ'' would have been 19.8 degrees in the example of Figure 7 while it would have been only 3.4 degrees in the example of Figure 9, the difference being due largely to the greater conducting angle $(\theta''-\theta')$ in the latter example. It should be clearly understood that the assumption of negligible exponential terms in equation 23 does not say that the exponential terms in the current equations of Mode II or III of Table I are negligible, but it says rather that the values these terms have in the current equations have become very small by the time the current ceases flowing.

Another fact which will be very useful in discussing wave shapes is that it is impossible by induction to produce a steady-state direct current in a circuit which contains no rectifying elements. In other words, the average value of the current flowing in the control circuit of Figure 1 must be equal to E_d/R_d when the amplifier is in a steady-state condition. It can be shown throughout the normal control range that this statement is equivalent to the conditions set forth in equation 24. This average value condition will be used to facilitate visualization in the following discussion.

As an introduction to the practical problem of parameter variation, the effect of varying E_d will be examined in a circuit which has "ideal magnetic amplifier cores," that is, cores in which the slope m of the magnetization curves of Figure 2 is zero, and no external linear inductances. Under these conditions, $L_a'=0$ and $L_d'=0$, the current equation in Table I, Mode II, becomes

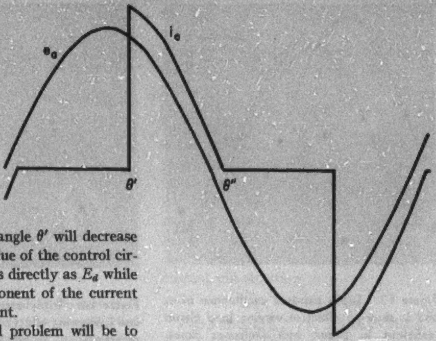
$$i_a=\frac{E_a}{R_a+R_d'}\sin(\omega t+\theta')+\frac{N_a}{N_d}\,\frac{E_d}{R_a+R_d'}$$

while equation 25 becomes

$$\sin\theta''=-\frac{N_a}{N_d}\,\frac{E_d}{E_a}$$

The circuit currents thus will jump instantaneously from zero to a finite value at θ', and then will follow a displaced sinusoid until the current reaches zero at θ''; Figure 10. As E_d is increased $\sin\theta''$ becomes more negative so θ'' increases; however, if $E_a>>E_d$ the increase in θ'' will be small, although the change in E_d may be large. The angle θ' will decrease since the average value of the control circuit current increases directly as E_d while the sinusoidal component of the current wave remains constant.

Figure 10. A-c supply voltage and load circuit current when $L_a'=L_d'=0$



The first practical problem will be to examine the effect of increasing E_d throughout its normal control range in the circuit of Figure 1 when $L_a=0$, $L_d=0$, and the magnetization constraints of the cores are as in Figure 2. If the slope m is small, which is the usual case with cores such as Deltamax, then the current waves will be quite similar to those in Figure 10, differing significantly in only one point. The current, instead of instantaneously jumping from zero to a finite value at θ', now rises smoothly, although quite rapidly, immediately after θ'. This deviation from the vertical jump is brought about by the exponential terms in the current equations for Mode II of Table I. The value of these exponential terms decreases quite rapidly as ωt increases and consequently the current waves resemble those of Figure 10 quite closely a few degrees beyond θ'. As the control voltage E_d is increased θ'' increases and θ' decreases in accord with the argument presented in the preceding paragraph. Figure 11, illustrating these variations, is a triply exposed oscillogram showing the a-c line voltage and the load circuit current for three different values of control voltage.

Figure 12 is an oscillogram triply exposed for the same three values of E_d as in Figure 11. In this case a linear inductance had been added to both the load and control circuits to give the complete circuit of Figure 1. The most significant difference between the current waves in the two pictures is that in the present case the value of the sum of the exponential terms in Modes II and III of Table I de-

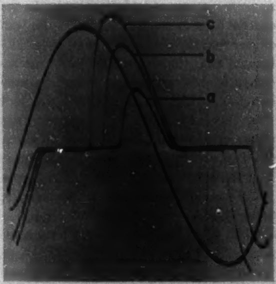


Figure 11. Triply exposed oscillogram of e_a and i_a showing effect of varying control circuit voltage in circuit with Deltamax cores. $E_a=78.5$ volts, $R_a=108.9$ ohms, $L=0$, $N_a=1{,}000$ turns, $R_d=109.8$ ohms, $L_d=0$, $N_d=1{,}000$ turns for

a. $E_d=3.8$ volts
b. $E_d=9.2$ volts
c. $E_d=15.2$ volts

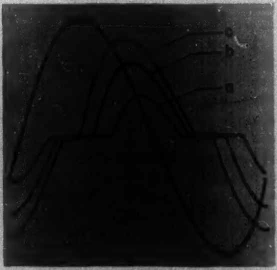


Figure 12. Triply exposed oscillogram of e_a and i_a showing effect of varying control circuit voltage in circuit with Deltamax cores. $E_a=78.5$ volts, $R_a=108.9$ ohms, $L_a=0.221$ henry, $N_a=1{,}000$ turns, $R_d=109.8$ ohms, $L_d=0.170$ henry, $N_d=1{,}000$ turns for

a. $E_d=3.8$ volts
b. $E_d=9.5$ volts
c. $E_d=15.2$ volts

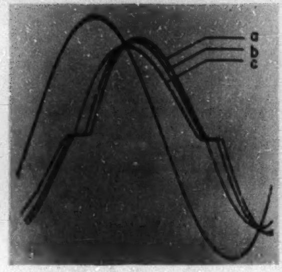


Figure 13. Triply exposed oscillogram of e_a and i_a showing effect of varying load circuit resistance in circuit with Deltamax cores. $E_a=78.5$ volts, $L_a=0.221$ henry, $N_a=1{,}000$ turns, $E_d=15.2$ volts, $R_d=109.8$ ohms, $L_d=0.170$ henry, $N_d=1{,}000$ turns for

- a. $R_a=108.9$ ohms
- b. $R_a=144.8$ ohms
- c. $R_a=204.8$ ohms

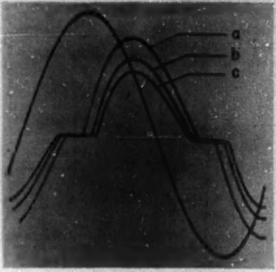


Figure 14. Triply exposed oscillogram of e_a and i_a showing effect of varying control circuit resistance in circuit with Deltamax cores. $E_a=78.5$ volts, $R_a=108.9$ ohms, $L_a=0.221$ henry, $N_a=1{,}000$ turns, $E_d=15.2$ volts, $L_d=0.170$ henry, $N_d=1{,}000$ turns for

- a. $R_d=109.8$ ohms
- b. $R_d=158.0$ ohms
- c. $R_d=206.0$ ohms

creases much more slowly as ωt increases than in the case of Figure 11, and consequently the current increases from zero at θ' much more slowly. Other differences can be seen by examining equation 25. The first factor on the right-hand side of this equation is greater in the present case than when $L_a=L_d=0$. Consequently, the difference in θ'''s for two values of E_d is greater in the present case than the corresponding difference in the case of Figure 11. Also the angle θ'' for a given E_d is itself greater in the present case than in the case of Figure 11, because both α and the factor on the right-hand side of equation 25 are greater in the present case.

The next problem will be to examine the effect of increasing the load circuit resistance R_a in the circuit of Figure 1. Examination of equation 25 shows that θ'' decreases as R_a increases since both α and the first factor on the right-hand side of the equation decrease. Increasing R_a also decreases the coefficients of the sinusoidal and constant displacement terms in the current equations of Mode II. The latter, coupled with the fact that the area under the current waves must remain constant as R_a increases, requires that the angle θ' decrease as R_a increases. Illustrations of this case are given in Figure 13.

The next problem will be to examine the effect of increasing the control circuit resistance R_d. Examination of equation 25 and of the current equations for Mode II shows that the control and load circuit resistances appear only as the sum R_a+R_d'. Consequently, increasing R_d will have exactly the same effect on the equations being discussed as increasing R_a. The one difference, which effects wave shape but which does not appear in these equations, is that increasing R_d decreases the average current which must flow in the control circuit. This decrease requires the angle θ' to increase as R_d increases. A triply exposed oscillogram showing this variation is given in Figure 14.

Next the effect of varying L_a will be examined. Increasing L_a increases both α and the first factor on the right-hand side of equation 25. Consequently, the angle θ'' increases as L_a increases. In the current equations for Mode II the coefficient of the sinusoidal component decreases as L_a increases while the time required for the exponential terms to decay increases. These factors, plus the fact that the area under the current waves must remain constant, combine to cause the angle θ' to decrease as L_a increases. This situation is illustrated in Figure 15(A) and (B), and will continue until the upper limit of the normal control range is reached when the current flows for a complete half-cycle, $\theta''=\theta'+180$ degrees. When L_a has such a value the flux in core I at θ'' will be Φ_s and the core will be proceeding toward unsaturation while the flux in core II will be $-\Phi_s$ and this core will be proceeding toward saturation. Further increase in L_a will cause both cores to be saturated simultaneously throughout a portion of the a-c supply voltage cycle. The equations which apply during this portion are given in Mode IV of Table I, and it can be seen that the control and load circuit currents will not have the same shape. The control circuit current will be an exponential increase, while the load circuit current will be the sum of an exponential and a sinusoid. This period of simultaneous saturation will occur immediately before the load circuit current passes through zero, and will continue until core I becomes unsaturated, at which time operation will begin in Mode III. The order of modes will thus be IV, III, IV, II, and repeat. The load circuit current wave for this case is shown in Figure 15(C).

Finally the effect of varying L_d will be studied. Examination of equation 25 and of the current equations in Modes II and III shows that the control circuit linear inductance appears only in the following sum combination $L_a'+L_d'$ or

$$4mN_a^2+\left(L_a+\frac{N_a^2}{N_d^2}\,L_d\right)$$

Furthermore, the average value of the control circuit current is unchanged by changing L_d. These facts show that throughout the normal control range of the amplifier the effect on current wave shape of varying L_d is exactly the same as varying L_a, Figure 16(A) and (B). Although the current waves are affected in the same way by changes in L_a and L_d the flux waves are not. This can be seen by noticing that L_a' and L_d' do not always appear as a sum in the equation for ϕ_{II} in Mode II or in the equation for ϕ_I in Mode III. Thus, although observation of the current waves alone might indicate

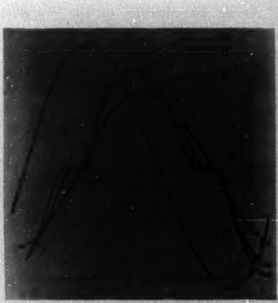


Figure 15. Triply exposed oscillogram of e_a and i_a showing effect of varying load circuit inductance in circuit with Deltamax cores. $E_a=78.5$ volts, $R_a=108.9$ ohms, $N_a=1{,}000$ turns, $E_d=15.2$ volts, $L_d=0$, $N_d=1{,}000$ turns for

- a. $L_a=0.221$ henry
- b. $L_a=0.391$ henry
- c. $L_a=0.817$ henry

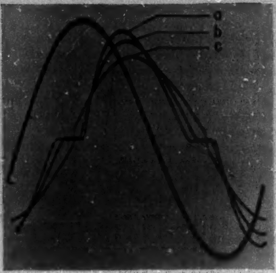


Figure 16. Triply exposed oscillogram of e_a and i_a showing effect of varying control circuit inductance in circuit with Deltamax cores. $E_a=78.5$ volts, $R_a=108.9$ ohms, $L_a=0$, $N_a=1{,}000$ turns, $E_d=15.2$ volts, $R_d=109.8$ ohms, $N_d=1{,}000$ turns for

a. $L_d=0.221$ henry
b. $L_d=0.391$ henry
c. $L_d=0.817$ henry

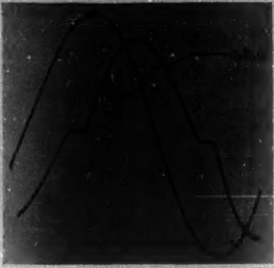


Figure 17. Triply exposed oscillogram of e_a and i_a showing effect of interchanging load and control circuit inductances in circuit with Deltamax cores. $E_a=78.5$ volts, $R_a=108.9$ ohms, $N_a=1{,}000$ turns, $E_d=15.2$ volts, $R_d=109.8$ ohms, $N_d=1{,}000$ turns for

a. $L_a=0.391$ henry and $L_d=0$
b. $L_a=0$ and $L_d=0.391$ henry
c. $L_a=0.221$ henry and $L_d=0.170$ henry

that changes in L_a and L_d have identical effects on amplifier operation, increasing L_d produces entirely different operating conditions from those which are produced by increasing L_a. This may be seen by comparing, in Figures 15(C) and 16(C) the load circuit current waves for values of L_a and L_d greater than the critical values which produce 180-degree current flow. Increasing L_d beyond its critical value eliminates operation in Mode I, the sequence of modes being Mode II, Mode III, and repeat.

Figure 17 is a triply exposed oscillogram verifying one of the results discussed in the preceding paragraph; namely, throughout the normal control range, the current wave shape depends upon the sum of the external linear inductances in the load and the control circuit and is independent of the division of the inductance between the two circuits.

Concluding Remarks

To examine the operating conditions of the series-connected magnetic amplifier when linear inductances are inserted in the load or control circuit, a mathematical model, determined by the four assumptions under the section entitled "Analysis," was established. The accuracy of the model was examined by quantitatively comparing the instantaneous values of the calculated and experimental current waves which exist in steady-state and in transient conditions. The mathematical model was used to explain the changes in the current wave shapes which occur when the control circuit resistance and inductance or the load circuit resistance and inductance are varied. It was shown that over a wide operating range the steady-state current waves depend upon the sum of the load and control circuit inductances and not upon the division of the inductances between the two circuits.

In discussing and calculating the transient current wave shapes the control voltage was considered to be a step increase wherein the voltage suddenly jumped from one value to another. Thus, the solutions in Table I and the methods described are obviously directly applicable where it is desired to find the current wave shapes in cases where the control voltage may be expressed as a sum of step functions, such as rectangular pulses, and so forth.

To obtain accurate quantitative prediction of the transient or steady-state wave shapes using the equations of Table I, it is necessary that the mathematical model, determined by the four assumptions under the section on "Analysis," closely match the physical system under consideration. In the example given under "Experimental Verification," this match was quite good. Nevertheless, examination of Figures 6, 7, and 8 shows two places where assumption 4 of the mathematical model does not agree too closely with the actual magnetization constraint of the physical system. First, the actual magnetization constraint is not a single-valued function but rather a multivalued function. Second, the slope of the saturated portion of the actual magnetization constraint is not constant, but rather decreases slightly as the magnetomotive force on the core increases. The former accounts for the fact that the load circuit current is not zero during Mode I when both cores are unsaturated, while the latter explains the slower increase and the slower decrease of the current waves which are evidenced by the rounding of the wave in the region of small current. Application of the methods presented herein to this more general problem and to more advanced circuitry is the subject of current investigation.

Appendix I. A Steady-State Derivation

Throughout the normal control range of the series-connected magnetic amplifier the steady-state fluxes and θ's are as shown in Figure 6 or

$$\Delta\phi_{\mathrm{I}} \text{ from } \theta'' \text{ to } \theta''' = \Delta\phi_{\mathrm{II}} \text{ from } \theta' \text{ to} \theta'' + \Delta\phi_{\mathrm{II}} \text{ from } \theta'' \text{ to } \theta''' \quad (26)$$

where Δ stands for "the change in." An expression for the first term in this equation can be obtained by taking the difference between the ϕ_{I} equations in Table I, Mode I, when $\theta=\theta''$, $\omega t=\theta'''-\theta''=\theta'-\theta''+\pi$ and when $\theta=\theta''$, $\omega t=0$.

$$\Delta\phi_{\mathrm{I}} \text{ from } \theta'' \text{ to } \theta''' = \frac{E_a}{\omega 2N_a}[\cos\theta''-\cos\theta''']+\frac{E_d(\theta'''-\theta'')}{\omega 2N_d}$$

Similarly, the last term in this equation may be written as

$$\Delta\phi_{\mathrm{II}} \text{ from } \theta'' \text{ to } \theta''' = \frac{E_a}{\omega 2N_a}[\cos\theta''-\cos\theta''']-\frac{E_d(\theta'''-\theta'')}{\omega 2N_d}$$

An expression for the first term on the right-hand side of the equation may be obtained by taking the difference between the ϕ_{II} equations in Table I, Mode II, when $\theta=\theta'$, $\omega t=\theta''-\theta'$, and when $\theta=\theta'$, $\omega t=0$.

$$\begin{aligned}\Delta\phi_{\mathrm{II}} \text{ from } \theta' \text{ to } \theta'' ={}& \frac{E_a}{\omega N_a}\frac{R_d'}{R_a+R_d'}\cos\theta'- \\ & \frac{E_a}{\omega N_a}\sqrt{\frac{(R_d')^2+(\omega L_d')^2}{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}}\times \\ & \cos(\theta''-\alpha+\beta)-\frac{1}{N_a}\frac{L_d'R_a-L_a'R_d'}{R_a+R_d'}\times \\ & \Bigg\{\frac{E_a}{\sqrt{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}}\times \\ & \sin(\theta'-\alpha)\times\exp\left[-\frac{(R_a+R_d')(\theta''-\theta')}{\omega(L_a'+L_d')}\right]- \\ & \frac{N_a}{N_d}\frac{E_d}{R_a+R_d'}\times \\ & \left[\!\left[1-\exp\left[-\frac{(R_a+R_d')(\theta''-\theta')}{\omega(L_a'+L_d')}\right]\right]\!\right]\Bigg\}- \\ & \frac{E_d}{\omega N_d}\frac{R_a(\theta''-\theta')}{R_a+R_d'}\end{aligned}$$

If these values are substituted in equation 26 and if then the equation is multiplied by $\omega N_a(R_a+R_d')/\boldsymbol{E}_aR_d'$, the result may be written as

$$\frac{N_a}{N_d}\frac{E_d}{\boldsymbol{E}_a}\left[\theta'-\theta''+\frac{R_a+R_d'}{R_d'}\pi\right]$$
$$=\frac{1}{\sqrt{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}}\times$$
$$\Bigg\{\sqrt{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}\times$$
$$\cos\theta'-(R_a+R_d')\cos(\theta''-\alpha)+$$
$$\omega L_d'(1+R_a/R_d')\sin(\theta''-\alpha)+$$
$$(\omega L_a'-\omega L_d'R_a/R_d')\Bigg[\!\!\Bigg[\sin(\theta'-\alpha)\times$$
$$\exp\left[-\frac{(R_a+R_d')(\theta''-\theta')}{\omega(L_a'+L_d')}\right]-$$
$$\frac{N_a}{N_d}\frac{E_d}{\boldsymbol{E}_a}\frac{\sqrt{(R_a+R_d')^2+\omega^2(L_a'+L_d')^2}}{R_a+R_d'}\times$$
$$\Bigg[\!\!\Bigg[1-\exp\left[-\frac{(R_a+R_d')(\theta''-\theta')}{\omega(L_a'+L_d')}\right]\Bigg]\!\!\Bigg]\Bigg]\!\!\Bigg]\Bigg\} \tag{27}$$

This result is true only in steady state since its formulation depended on the fact that the flux in core II at $\theta'''=\theta'+\pi$ must be equal in magnitude but opposite in sign to the flux in core I at θ'. Equation 23, however, is a statement only of the fact that, if the angle at which current begins to flow in an alternating voltage cycle is known, then the angle at which the current returns to zero is determined uniquely. This latter relation is true in the general transient case and consequently it may be used in the special case of steady state. Substituting equation 23 in equation 27 and combining terms yields

$$\cos\theta'-\frac{N_a}{N_d}\frac{E_d}{\boldsymbol{E}_a}\theta'=\cos\theta''-$$
$$\frac{N_a}{N_d}\frac{E_d}{\boldsymbol{E}_a}\theta''+\frac{R_a+R_d'}{R_d'}\frac{N_a}{N_d}\frac{E_d}{\boldsymbol{E}_a}\pi \tag{24}$$

Appendix II. List of Symbols

$\boldsymbol{E}_a$ = peak value of sinusoidal a-c supply voltage, volts
E_d = d-c control voltage, volts
i_a = instantaneous value of load circuit current, amperes
$i_a(0)$ = instantaneous value of load circuit current at θ, amperes
i_d = instantaneous value of control circuit current, amperes
$i_d(0)$ = instantaneous value of control circuit current at θ, amperes
L_a = external linear inductance in load circuit, henrys
$L_a' = L_a+2mN_a^2$, henrys
L_d = external linear inductance in control circuit, henrys
$L_d' = (N_a/N_d)^2L_d+2mN_a^2$, henrys
m = slope of saturated segment of magnetization curve, webers per ampere-turn
N_a = number of turns per load circuit reactor coil
N_d = number of turns per control circuit reactor coil
R_a = total resistance of load circuit, ohms
R_d = total resistance of control circuit, ohms
$R_d' = (N_a/N_d)^2R_d$, ohms
t = time measured from a discontinuity of E_d, ϕ_{I}, or ϕ_{II}, seconds
t' = time measured from zero in a-c supply voltage wave, seconds
$\alpha = \tan^{-1}[\omega(L_a'+L_d')/(R_a+R_d')]$
$\beta = \tan^{-1}[\omega L_d'/R_d']$
$\gamma = \tan^{-1}[\omega L_a'/R_a]$
θ = angle in a-c supply voltage wave at which a discontinuity in either E_d, ϕ_{I}, or ϕ_{II} occurs, radians or degrees
$\theta_1, \theta_2, \ldots$ = angles in a-c supply voltage wave at which first, second, ... discontinuity occurs, radians or degrees
θ' = angle in positive half-cycle of a-c supply voltage wave at which current begins flowing in steady state, radians or degrees
θ'' = angle in negative half-cycle of a-c supply voltage wave at which current ceases flowing in steady state, radians or degrees
ϕ_{I} = instantaneous value of the flux in core I, webers
$\Delta\phi_{\text{I}}$ = change in flux in core I, webers
$\phi_{\text{I}}(0)$ = value of flux in core I at θ, webers
ϕ_{II} = instantaneous value of the flux in core II, webers
$\Delta\phi_{\text{II}}$ = change in flux in core II, webers
$\phi_{\text{II}}(0)$ = value of flux in core II at θ, webers
Φ_s = saturation flux, webers
ω = angular frequency of a-c supply voltage, radians per second

References

1. Response Time of Magnetic Amplifiers, **E. L. Harder, W. F. Horton.** *AIEE Transactions*, volume 69, part II, 1950, pages 1130–41.

2. An Analysis of Transients and Feedback in Magnetic Amplifiers, **Walter C. Johnson, Frank W. Latson.** *AIEE Transactions*, volume 69, part I, 1950, pages 604–12.

3. On the Mechanics of Magnetic Amplifier Operation, **Robert A. Ramey.** *AIEE Transactions*, volume 70, part II, 1951, pages 1214–23.

4. The Theory of the Current Transductor and its Application in the Aluminum Industry, **T. R. Specht, R. N. Wagner.** *AIEE Transactions*, volume 69, part I, 1950, pages 441–52.

5. Some Fundamentals of D-C Controlled Reactors with Resistive Load, **H. F. Storm.** *AIEE Transactions*, volume 68, part I, 1949, pages 188–94.

6. Series-Connected Saturable Reactor with Control Source of Comparatively High Impedance, **H. F. Storm.** *AIEE Transactions*, volume 69, part II, 1950, pages 1299–1309.

7. Series-Connected Saturable Reactor with Control Source of Comparatively Low Impedance, **H. F. Storm.** *AIEE Transactions*, volume 69, part II, 1950, pages 756–65.

8. Transient Response of Saturable Reactors with Resistive Load, **H. F. Storm.** *AIEE Transactions*, volume 70, part I, 1951, pages 95–102.

9. An Analysis of Transients in Magnetic Amplifiers, **D. W. Ver Planck, L. A. Finzi, D. C. Beaumariage.** *AIEE Transactions*, volume 69, part I, 1950, pages 498–503.

10. Mathematical Analysis of Non-Linear Circuits, **A. Boyajian.** *General Electric Review* (Schenectady, N. Y.), volume 34, 1931.

11. Properties of Deltamax 4-79 Mo-Permalloy, Supermalloy. *Bulletin TC-100*, The Arnold Engineering Company (Marengo, Ill.), March 15, 1951, page 9.

No Discussion

Automatic Regulation of Metallic Rectifiers by Magnetic Control

D. H. SMITH
MEMBER AIEE

Synopsis: Regulated rectifiers have numerous applications in the Bell System and recent years have seen increased use of selenium cells as the rectifying elements in these units. In large power plants, regulation has been achieved using reference standards and electronic amplifiers in checkback circuits. In plants having less than 50 amperes d-c charging capacity, however, it has been possible to use magnetic amplifier techniques to secure ±1 per cent regulation, with an additional advantage that periodically renewable circuit elements have been eliminated in the regulating circuit.

AT the conclusion of World War II, high-grade selenium rectifier stacks became available in quantity and were used in the Bell System as rectifying elements in large regulated battery chargers of 100- and 200-ampere capacity.

These large rectifiers were capable of maintaining accurate floating and charging potentials at the storage battery, using electronic means for automatically adjusting the rectifier output to compensate for load current and line voltage changes.[1]

Experience with these large rectifiers indicated that voltage regulation did a very satisfactory job in maintaining a full charge in the battery at all times. In addition, current regulation provided automatic protection from rectifier overload following primary power failure.

Quite logically, it was then apparent that a need existed for inexpensive regulated rectifiers of various current capacities from 1 ampere to 9 amperes for charging and floating the batteries associated with Private Branch Exchanges (PBX) and other installations where the battery, although much smaller than that used in central offices, nevertheless represented a considerable proportion of the total cost of the power plant.

A comprehensive study revealed that the majority of PBX installations could utilize one of three standard sizes in the power plant. The general requirements for this line of rectifiers are summarized in Table I.

The requirements for small charging sets are, in general, less stringent than those having higher ratings. First, the small plants are mostly used with equipments where voltage tolerances are not limited by the operating requirements of complex automatic circuits. Second, the charging set is usually not required to be switched in and out of the power plant as load demand varies, a technique quite advantageous in larger plants. Finally, load demand in a small plant builds up or decays slowly as a rule, and thus a fast dynamic regulation characteristic is not a principle design criterion, although the speed of response must be optimum if plant stability is to be maintained.

The smaller charging sets, however, must have features permitting automatic recharging of the storage battery after power failure, which involves designing the rectifier to operate on a constant voltage basis up to full rated load, and on a constant current basis at loads above rated full output. This is necessary because the battery reserve will sometimes be discharged to approximately 1.75 volts per cell during power failures, and when primary power is then reapplied to the unit, the discharged battery will have very low counter electromotive force, tending to severely overload the rectifier unless protection is built-in to the circuit. These rectifiers must therefore be capable of supplying enough charging current for the normal equipment load, in addition to furnishing sufficient power for recharging the battery in a reasonable amount of time after heavy discharge.

A review of voltage and current regulation methods revealed that the majority available at that time involved one or more vacuum tubes or other renewable components, and these were considered undesirable from a maintenance standpoint. The objective of developing long life telephone apparatus requiring little or no attention and having component parts which ordinarily do not require replacement during the life of the equipment is highly justifiable in power supply design, and development effort was therefore concentrated upon static devices. Magnetic-amplifier techniques had been enjoying increased popularity in this country, and preliminary studies indicated that such techniques were readily applicable to these problems.

The end result of our studies was a rectifier unit rated at 24 volts, 9 amperes direct current and this rectifier served as a prototype for the other rectifiers in the line. The circuit of the rectifier is shown in Figure 1.

Description of Operation

The portion of the circuit shown in Figure 1 to the left of the broken line is called the line voltage regulator, and serves to reduce ±8 per cent input voltage changes to about ±1 per cent or better. This portion of the circuit involves ferro-resonant techniques, and is outside the scope of the present paper.

The operation of the load regulating circuit shown in Figure 1 to the right of the broken line can be explained on an impedance basis, neglecting power factor

---

Paper 52-102, recommended by the AIEE Metallic Rectifiers Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25. Manuscript submitted October 25, 1951; made available for printing December 17, 1951.

D. H. SMITH is with Bell Telephone Laboratories, Inc., New York, N. Y.

The fundamental principles of the regulating circuit were worked out by J. A. Potter, and R. R. Gay was responsible for the equipment design of these rectifiers.

Table I. Rectifier Requirements

Feature	Limits
A-c input service	115±8% volts, 60±1% cycles
Overall d-c regulation at the battery	Nominal floating voltage for 10, 11, or 12 cells±2%
Ambient temperature range	Regulation accuracy to be maintained over range of 4 to 43° C
Output noise at battery	35 decibels (adjusted for F1A line weighting)
Overload	130–150% continuous
Expected life	20–30 years

and waveform effects, and it will be so explained for the sake of simplicity.

Consider the relation of the impedance of the primary winding of saturable transformer (T1) and the impedance of ballast reactor (L1). The total input voltage appears across these two windings in series, and the proportion of this voltage across the primary is dependent upon the degree of saturation in the core of (T1).

The d-c windings on (T1) are so arranged that their fluxes oppose in the core and cancel each other at full load. The shunt d-c winding is connected directly to the battery through a resistor network and a relay contact. This winding contributes a relatively constant amount of flux to the core. The series d-c winding carries the load current and consequently contributes an increasing amount of flux to the core as the output current is increased to full load. The net saturation in the core is thus high at no load and almost zero at full load.

With high saturation in the core, the impedance of the primary a-c winding is fairly low, and a certain proportion of the input voltage is developed across it and by transformer action, across the secondary. The a-c voltage is rectified and applied to the battery and load. As the load current demand increases, the impedance of the primary is made to increase in proportion to the losses in the system by the change in core saturation. At full load, the ampere turns of the two d-c windings are equal and opposite, and the mutual cancellation of their fluxes results in maximum primary impedance, maximum voltage drop across the primary winding and maximum output voltage.

If for any reason the output current increases above full load, the series d-c winding begins to re-saturate the core, lowering the impedance of the a-c primary and causing the output voltage to decrease. This results eventually in less ampere turns in the d-c shunt winding and tends to further saturate the core and further reduce the output voltage. Overload protection is thus obtained by an automatic transition from constant voltage regulation to constant current regulation at output currents greater than full load.

By means of taps on the series d-c winding to adjust the compounding slope, and taps on the line regulator to adjust the degree of line voltage correction, the battery voltage can be maintained at the recommended floating voltage ±1 per cent.

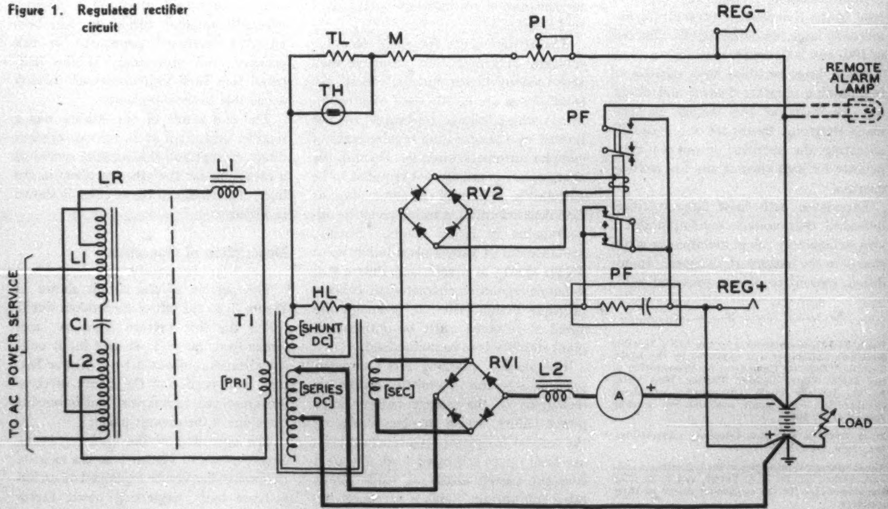
Ambient Temperature Errors

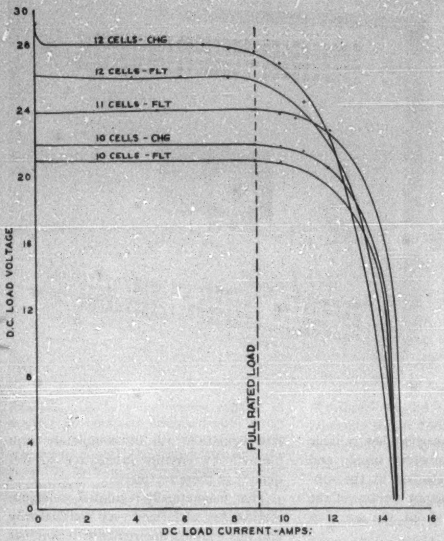
There are several major sources of ambient temperature error in the circuit including the temperature coefficient of the copper in the shunt d-c winding and the temperature coefficient of the forward drop of selenium rectifier (RV1). Both act in a positive direction, that is, cause the output voltage of the rectifier to rise with an increase in ambient temperature. The total ambient error is corrected with a thermistor or negative temperature coefficient resistor designated (TH) in the circuit, inserted in series with the shunt d-c winding. The potentiometer (P1) serves to adjust the regulated voltage over the range of 21 to 28 volts allowing 10, 11, or 12 cells of low-gravity storage battery to be floated at 2.15 volts per cell or charged at 2.33 volts per cell. A relay contact in series with the regulating leads prevents battery discharge through the shunt d-c winding in the event of power failure.

Personnel Hazard

Resistance (HL) was added to the circuit when it was found that 120-cycle harmonic voltage in the series d-c winding of a few millivolts magnitude was stepped up by the turns ratio of the two

Figure 1. Regulated rectifier circuit





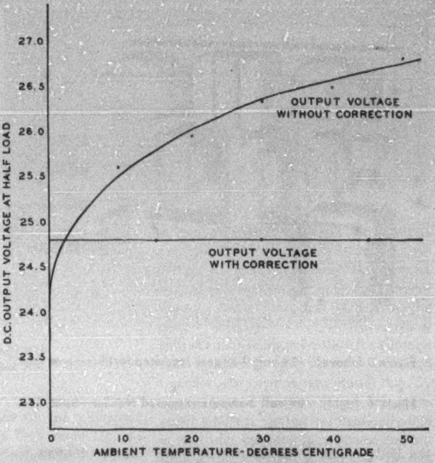


Figure 2 (left). Regulated rectifier, 24 volts, 9 amperes—load regulation curves

Figure 3 (above). 24-volt 3-ampere regulated rectifier—correction for ambient temperature changes

d-c windings to a hazardous level. Under normal operating conditions, the conductance of the resistor network (TH, TL, M, and P1) and the low impedance battery was sufficient to reduce this a-c voltage to almost zero, but if a fuse, a resistor or the relay contact were inadvertently opened, a potential of about 400 volts was developed across the open component. By loading this winding down with the (HL) resistor, the a-c potential was reduced to a safe value. The induced harmonic voltage could probably have been reduced further by connecting the two halves of the a-c primary windings in parallel, but this would have decreased the speed of response of the regulating action.

It is of interest to note that the need for a cover switch has been circumvented in this line of rectifiers by two methods. The first, discussed in the preceding paragraph, was to reduce operating and trouble voltage on components to a value of 150 volts or less to ground. Where this could not be done, the second method was employed and is demonstrated in the 3- and 9-ampere sizes on the line regulator panel. A ferroresonant circuit is employed to effect line voltage regulation, and since it was impracticable to reduce the operating voltage on the components to 150 volts or lower, that portion of the complete circuit is covered by a red plastic grid which permits test picks to be introduced through the grid squares but protects the operator from inadvertent contact with high voltage. In addition, a decal warning sticker appears on the line regulator panel.

Performance

Production samples of this type of regulated rectifier have performed satisfactorily in service and have met all design requirements. Typical regulation curves for a 24-volt 9-ampere rectifier are shown in Figure 2. Figure 3 is a curve showing the degree of temperature correction necessary in the 24-volt 3-ampere model which was considerably more than that for the 9-ampere unit. Figure 4 is a front view of the 24-volt 9-ampere rectifier with the cover removed and Figure 5 is a rear view of the same unit. The 9-ampere unit is used in the power plant for

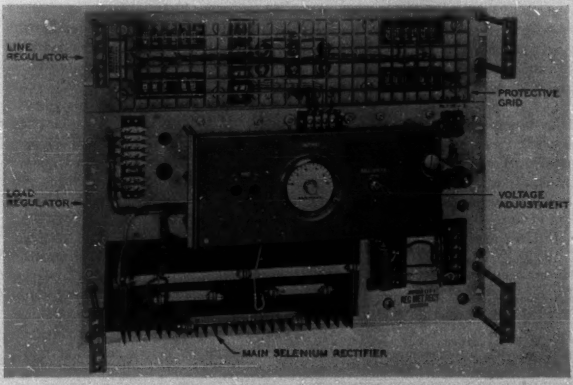


Figure 4. 24-volt 9-ampere regulated rectifier—front view, cover removed

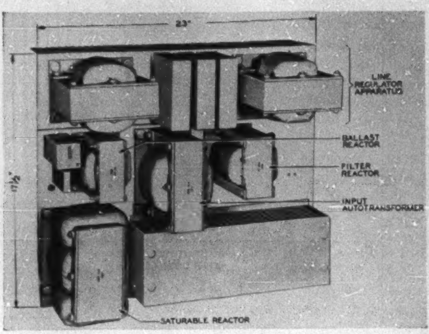


Figure 5 (above). 24-volt 9-ampere regulated rectifier—rear view

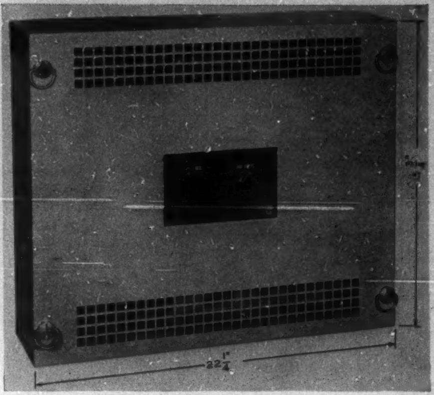


Figure 6 (right). 24-volt 3-ampere regulated rectifier—front view

the Bell System Transcontinental Radio Relay stations, and was designed for mounting in relay racks or duct-type framework. Figure 6 is a front view of the 3-ampere rectifier used in a voice-frequency repeater power plant. The 24-volt 1-ampere rectifier has been developed but is not yet in production.

Field experience with these units has indicated that the ferroresonant type of line voltage regulation system may cause some difficulty when the rectifiers are required to operate from an a-c emergency power source such as an engine alternator. The frequency regulation of these emergency sources is somewhat less reliable than that of the commercial supply and since the circuit is resonant at the supply frequency, the output voltage of the rectifier may depart from the specified value.

It has been possible to apply saturable reactor techniques to the problem of line regulation and in cases where variable frequency power service is encountered, these techniques have been employed. The 24-volt, 1-ampere rectifier utilizes this method of line regulation. As magnetic amplifier techniques improve, it is possible that the ferroresonant type of line regulation will be completely supplanted by circuits insensitive to frequency in these rectifiers.

The magnetically-regulated selenium rectifier has performed very satisfactorily and has enabled maintenance techniques to be considerably simplified. The rectiers have been designed for a trouble-free life expectancy of 20-years and our experience to date indicates that this should be easily realizable.

Reference

1. METALLIC RECTIFIERS IN TELEPHONE POWER PLANTS, D. E. Trucksess. *AIEE Transactions*, volume 70, part II, 1951, pages 1464–67.

No Discussion

A High-Speed Direct-Scanning Facsimile System

C. R. DEIBERT
NONMEMBER AIEE

F. T. TURNER
NONMEMBER AIEE

R. H. SNIDER
NONMEMBER AIEE

Synopsis: A facsimile system capable of sending 80 complete standard size letters per hour is described. Direct scanning is employed at both the transmitter and the recorder, making the transmitted message immediately available at the receiving terminal. Servoloops around the 1,800-rpm scanning motors provide the synchronous precision required at the speeds involved. Phasing is accomplished by controlling the phase of the motor driving voltage at the recorder. The equipment is automatic and under the complete control of the sending operator.

FACSIMILE, although almost as old as Morse telegraphy, did not come into widespread use as a communication device until after World War I. This is perhaps the result of a series of circumstances. In the first place, the Phoenicians are said to have developed the alphabet which is a cleverly contrived code making possible the recording of all of the sounds of spoken language by means of comparatively few symbols. Next must be added the work of Morse and Baudot who converted these symbols, the letters of the alphabet, into an even simpler code whereby the written language could be transmitted by wire using a series of electric impulses. This makes possible the transmission of any intelligence which can be expressed in language by comparatively inexpensive and simple circuits. Facsimile, on the other hand, scans the copy in terms of the elemental area, which in design can be varied to meet the definition requirements of any particular application. Anything which can be put on paper, whether it be written, drawn, printed, or painted, can be transmitted with equal facility. This versatility, however, does not come free; it must be paid for either in terms of time or spectrum bandwidth. If all languages had developed as did that of the Chinese, facsimile might today be a more highly developed and common means of communication.

In time, so many channels were required for the simultaneous transmission of a large number of messages at the same time that means were developed for stacking numbers of these on a single facility. The ultimate to date is the development of the coaxial cable and the microwave relay system where a very large quantity of terminal equipment is required for the combining and extracting of the large number of separate channels which can be handled.

It is obvious, at least where record copy is involved (that is, intelligence on a piece of paper), that were these copies to be sent in time sequence at high speed less terminal equipment might be required. Also, if the high speed facility were simple enough to operate and maintain, although not as efficient of spectrum space as other means, conceivably it could be more economical to operate and maintain.

Thus we must reach the conclusion that, where facsimile is used, it is and must be an economical solution to a communication problem rather than a scientific possibility, for systems have already been developed capable of sending full page letters at the rate of 15 per second and this is probably by no means the ultimate if circumstances require higher speeds. High speed is, therefore, a solution only so long as it creates no new problem of an economically serious character, namely, the problem of gathering and distributing the material at the terminals.

The Western Union facsimile terminal described in this paper, Figure 1, is believed to be the highest speed equipment announced to date employing direct scanning at both the transmitter and recorder. The original piece of copy, so long as it is a sheet of paper with dimensions no greater than $8^1/_2$ by 15 inches, is placed directly into the transmitting cylinder and is scanned directly through that cylinder by optical means. At the recorder, the copy is reproduced directly onto a sheet of Western Union Teledeltos paper which is ejected automatically and is ready for delivery in a permanent dry form within a second or two after the last scanning line has been completed.

Terminal Components

A complete terminal consists of two transmitters operating alternately to keep the circuit in use continuously, a series of panels including amplifiers, power supplies, and control equipment, and two recorders. The second recorder is a stand-by unit in the sense that one recorder is used continuously until its roll of paper is exhausted at which time operation is switched to the stand-by unit.

The speed is such that a series of $8^1/_2$ by 11-inch letters can be sent in their entirety at the rate of one every 45 seconds or 80 per hour. Copy is scanned at the rate of $2^1/_8$ square inches per second or 1/4 lineal inch per second. In the telegraphers' language, this corresponds to 1,250 words per minute for pica type or 1,500 words per minute for elite type. Definition is determined by an $8^1/_3$-mil diameter elemental area.

The facsimile signal developed at the transmitting photocell covers the spectrum from a fraction of a cycle per second to 15.5 kc for equal horizontal and vertical definition. When modulating a car-

---

Paper 52-103, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, January 21–25, 1952. Manuscript submitted October 23, 1951; made available for printing December 11, 1951.

C. R. DEIBERT, F. T. TURNER, and R. H. SNIDER are all with The Western Union Telegraph Company, New York, N. Y.

The authors acknowledge their indebtedness to R. J. Wise and his staff for invaluable advice and assistance.

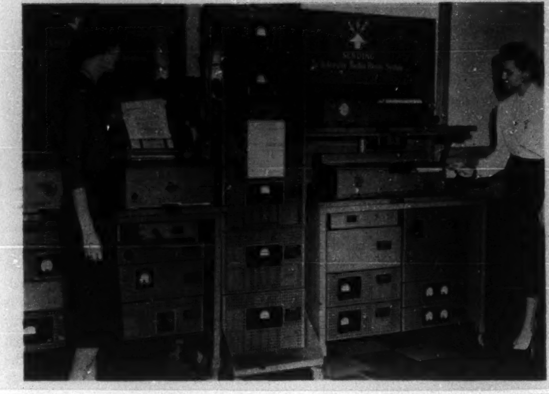


Figure 1. A complete high-speed facsimile terminal

rier on a double side-band basis, the bandwidth required is 31 kc. To date, this equipment has been operating on a double-side-band basis; however, single-side-band operation is envisaged for the future.

The functions of a facsimile system may be divided into three categories: scanning, synchronizing, and phasing. Scanning is the process of reading the original copy, point by point, at the transmitter and laying down a copy of that original at the recorder. Synchronizing is the process of keeping the scanning motors at both ends of the system running at the same speed during a transmission so that the intelligence laid down at the recorder shall fall in the same relative position as it occupied on the original. Phasing, which may be called framing, is the process of setting up the equipment so that the reproduced copy shall be reasonably well centered on the sheet.

Transmitter Scans at 1,800 Rpm

Scanning at the transmitter, Figure 2, is accomplished in the following manner: The copy is placed within a transparent cylinder which revolves at the rate of 1,800 rpm. A carriage mounting a lamp, optical system, and photocell travels along the length of the cylinder at a uniform rate, thus scanning the inside surface of the cylinder in a continuous spiral. The area under an objective lens is floodlighted by means of an incandescent lamp and condenser lens. The objective lens images the lighted area onto an opaque plate with an $8^{1}/_{2}$-mil dissecting aperture at its center. The light passing through the aperture falls on the cathode of a multiplier phototube causing current emission proportional to the reflectivity of the area being scanned. By amplification within the phototube, the emission current is transformed to a peak voltage swing of up to 50 volts in amplitude. This voltage swings from a fixed positive value for black or no reflectivity to a less positive value dependent upon the background reflectivity of the copy. A higher degree of versatility would be achieved if this voltage swing could be converted to one wherein the negative peak would be the reference and always zero. Consider for a moment that the preponderance of written, printed, or drawn intelligence is written in black or some dark color on a light background. If we always record from the background toward black, that is, record in various shades of gray toward black on our recording medium, we shall in general have better, cleaner-looking copy than if we record from black toward white making the background various shades of gray.

To achieve this end, although we al-

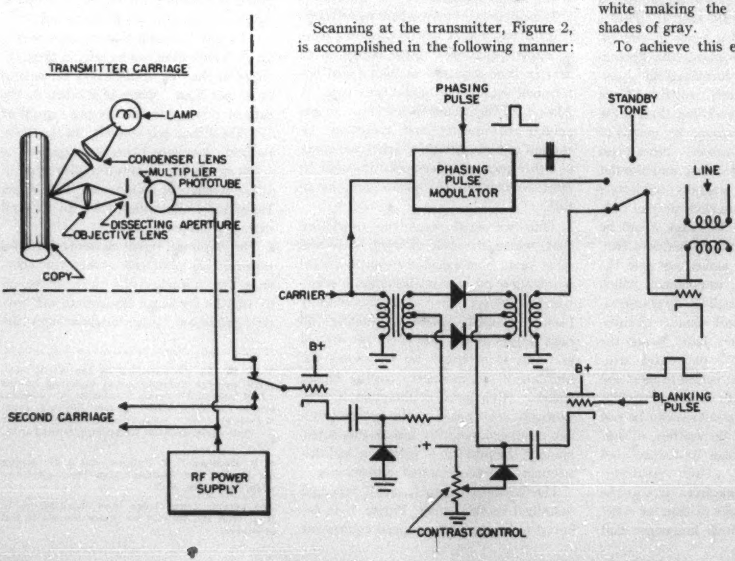


Figure 2. Transmitter scanning system

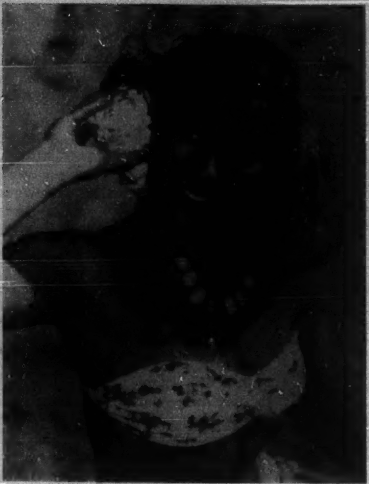


Figure 3 (left). High-speed facsimile reproduction of continuous-tone original

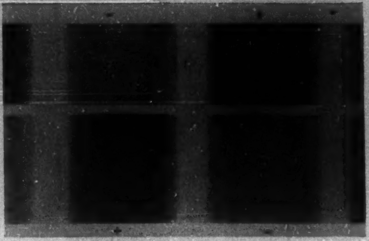


Figure 4 (right). Effect of polarity on the density of recording

ready have a d-c signal, a d-c restorer circuit has been added which always clamps the peak negative signal, regardless of its absolute value, to zero. The restorer was so designed that, when pictorial copy with a continuous tone scale is to be sent, the original tone scale is preserved so long as the copy has a margin at least 1/4 inch wide. From the restorer, the signal is fed to a conventional balanced modulator and thence through amplifiers to the line.

Background Signal Reduced

In the modulator, two other functions are performed. First, a bias is applied to the modulator to eliminate that signal which is due to paper texture itself. It was found that there was a wide variation in the signal derived from the surface of various types of paper. Accordingly, a bias control was added to eliminate all but a trace of this signal which, insofar as the recorder is concerned, is actually little more than noise. This resulted in an apparent increase in signal-to-noise ratio for even light marks.

The second additional function is to blank out the paper gap. The equipment is designed to handle variable widths of copy up to $8^1/_2$ inches. The inside circumference of the cylinder is $8^5/_8$ inches so that there is always a paper gap of 1/8 inch or more. If not blanked out, it would be recorded as black. The blanking is accomplished by biasing off the modulator completely during the time the paper gap is being scanned. This circuit too must include a d-c restorer because the signal is obtained from a plate circuit through a capacitor. Otherwise, the signal required to accomplish blanking would appear to vary with the length of the paper gap pulse.

At the recorder, scanning is accomplished by four styli mounted on a continuous belt and Western Union's direct recording Teledeltos paper. The distance between adjacent styli is equal to the inside circumference of the transmitter drum. The belt is driven by a motor identical to that used at the transmitting end. Here the styli continuously scan a fixed line; the paper is gradually fed by at the same rate as the carriage is advanced at the transmitter. The received signal is amplified and applied to the belt and styli. The current flowing from the styli through the Teledeltos paper to ground marks the paper, the density of the mark varying with the amplitude of the signal received from the transmitter. Although the recording characteristics of Teledeltos paper are not quite linear enough to give the best possible results insofar as the reproduction of photographs is concerned, no effort was made to minimize this defect in the first design. Figure 3 is an example of what can be accomplished.

The recording amplifier is a power amplifier capable of delivering up to 75 watts. This amplifier works on the modulated carrier directly. Because it was found that positive polarity extended the recording range somewhat toward black, a full-wave rectifier was included in the output to take advantage of this phenomenon; see Figure 4.

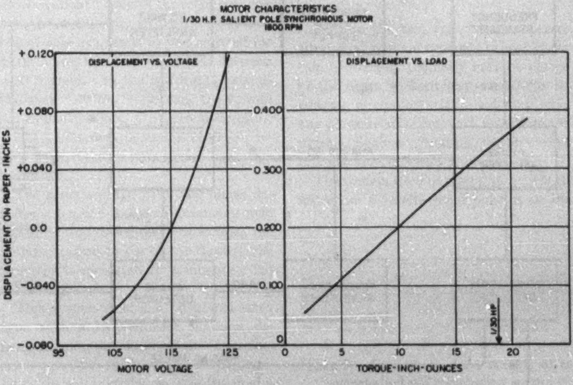


Figure 5. Motor voltage and load characteristics

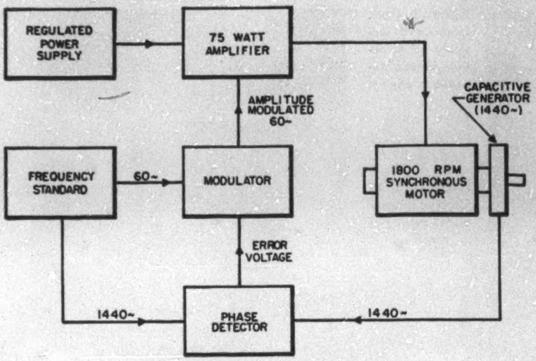

REGULATED POWER SUPPLY
75 WATT AMPLIFIER
AMPLITUDE MODULATED 60~
CAPACITIVE GENERATOR (1440~)
FREQUENCY STANDARD
60~
MODULATOR
1800 RPM SYNCHRONOUS MOTOR
ERROR VOLTAGE
1440~
PHASE DETECTOR
1440~

Figure 6. Transmitter drive system

Precision of Synchronization

Synchronization has been defined as the process of keeping the scanning motors at both ends of the system running at the same speed. This identity must be maintained to a very high degree of precision, the exact degree depending somewhat on the nature of the disturbance. For instance, a disturbance at either the transmitter or recorder, which shifts one with respect to the other by as little as 0.004 inch, is detected readily if it occurs during the scanning of a single line of type or less, whereas a larger disturbance, if evenly distributed over, say, 5 or 10 lines of type might go completely undetected.

The two important sources of disturbances are load changes and changes in motor power, be they voltage or frequency. Frequency differences and changes can be eliminated by driving the motors through amplifiers from frequency standards with accuracies of four parts per million or better. A relative difference between the two ends of the circuits of such an amplitude would result in a skew no greater than 1/16 inch for a transmission 14 inches long. Commercial frequency standards are readily available capable of giving such performance. From the curve of Figure 5 it will be noted that around 115 volts the deviation on the face of the facsimile copy is 0.008 inch for a change of 1 volt applied to the motor. Therefore, to maintain the equipment within the 0.004 inch previously mentioned, the output of the amplifiers at each end must remain relatively constant to within about 1/2 volt or ± 1/4 volt for each amplifier. To achieve this requirement as nearly as possible, the plate and bias supplies associated with the motor amplifiers and the frequency standard are of the regulated type. This leaves us with the load variations which can be minimized, but not to the required extent, by careful design. Even in the transmitter, where the load consists entirely of windage and bearing friction, the variation in the ball bearing load alone is enough to cause undesirable disturbances. In the recorder where, for good registration, stylus-restraining forces are required, variations in load become quite serious. The curve of Figure 5 indicates that to keep the system synchronous to 0.004 inch, the load variation of one terminal with respect to the other may be no greater than about 0.2 inch-ounce. If the tolerance is divided equally between the two ends of the system, each load must remain constant within 0.1 inch-ounce or within ± 1 per cent.

One method of minimizing the effect of load variations is to add a constant load of such proportions as to render the variations inconsequential. Since we are after an improvement of the order of about ten times, such an approach would require larger motors and larger amplifiers and would result in an unwieldy, inefficient, and uneconomical solution.

Experience with other equipment indicated that some form of servo loop was necessary to counteract those disturbances which could not be eliminated economically.

Flywheel Concept Applied

To fulfill this need, a device was developed which, in a sense, compares the

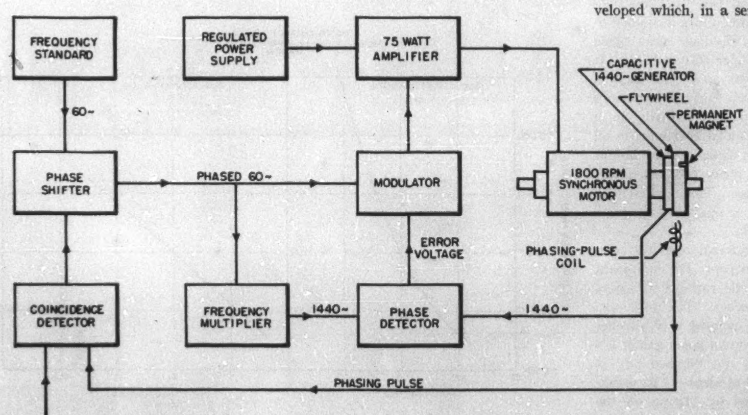

FREQUENCY STANDARD
REGULATED POWER SUPPLY
75 WATT AMPLIFIER
CAPACITIVE 1440~GENERATOR
FLYWHEEL
PERMANENT MAGNET
60~
PHASE SHIFTER
PHASED 60~
MODULATOR
1800 RPM SYNCHRONOUS MOTOR
ERROR VOLTAGE
PHASING-PULSE COIL
COINCIDENCE DETECTOR
FREQUENCY MULTIPLIER
1440~
PHASE DETECTOR
1440~
PHASING PULSE
PHASING PULSE FROM TRANSMITTER

Figure 7. Recorder drive and phasing systems

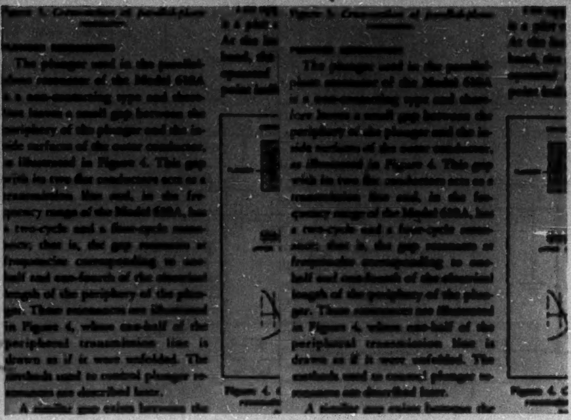


Figure 8. Comparison of transmissions with and without motor stabilization

instantaneous position of the rotating mass of the equipment with an ideal flywheel and develops from this comparison an error voltage proportional to the relative displacement of one with respect to the other. The error voltage in turn is applied in such a manner as to reduce the original displacement to negligible proportions.

In the equipment, the actual comparison is made between two alternating potentials; see Figure 6. One, the ideal, is derived from the frequency standard; the other, from a capacitive-type generator mounted on the motor shaft. A phase detector compares these two potentials and delivers an error signal which is proportional to the displacement of the rotor from its ideal position. The error voltage is applied to a modulator which regulates the amplitude of the standard frequency applied to the amplifier and thus to the motor to compensate for the disturbing influence.

The heart of the frequency standard is a 240-cycle temperature-compensated fork. This frequency is divided to 60 cycles for the motor circuits and multiplied to 1,440 cycles for the transmitter phase detector. These signals are fed through isolation amplifiers to the transmitter and recorder.

The phase detector used in the conventional circuit consisting of two transformers feeding a double diode. The output is, within limits, proportional to the cosine of the phase difference between the two applied voltages and the amplitude of the smaller of the two voltages. In this equipment, a frequency of 1,440 cycles was selected to apply to the phase detector to increase its sensitivity and eliminate the need for a stable d-c amplifier in its output. That the selection of a substantially higher frequency increases the sensitivity of the phase detector can be explained as follows: By designing the generator so that it developes 48 complete cycles per motor revolution, a displacement on the copy of 0.004 inch, which corresponds to 1/6 degree on the motor shaft, causes a phase shift of the generated tone of 8 degrees. Gain of this sort can be taken advantage of up to a point. If, during operation, disturbances are encountered which cause displacements just over 90 degrees of the generated tone, the motor and servo system will jump over to a new position which is about 1 cycle of the reference tone away from the original position. A system of this sort tends to restore itself for disturbances which shift the tone between ±90 degrees. In the quadrants extending from ±90 to 180 degrees, the system is unstable and jumps to a new position.

The tone generator itself is of the capacitive type because of its comparative simplicity.

The error voltage from the phase detector is applied to the modulator in such a manner that, for an increase in load, the voltage applied to the motor through the amplifier is increased to compensate for the load change.

This system is of the positional-servo type with a loop-transfer function dependent upon the motor sensitivity, in degrees per volt, the phase detector sensitivity, in volts per degree, and the modulator and amplifier gain. All systems of this type tend to hunt or oscillate as the loop gain is increased. The insertion of a proportional-plus-derivative network between the phase detector and the modulator allowed the loop transfer function to be increased to the point where satisfactory operation was attained. The transfer function achieved results in a stability with respect to small load variations which is equivalent to a 15 times increase in the running torque of the motor.

The motor drive system, Figure 7, at the recorder differs somewhat from that at the transmitter because of the necessity of phasing the recorder motor. The standard 1,440 cycles must be derived by multiplication from the phased 60 cycles as the reference tone as well as the motor must be phased properly to achieve satisfactory stabilization and accurate phasing.

Figure 8 compares transmissions of 9-point type with and without the servo-loops acting. The figure outline shows the extent of the displacement of one terminal with respect to the other. The accompanying distortion in the text is quite obvious. Figure 9 shows on an amplified scale the variation of an uncontrolled and a controlled motor and indicates the effectiveness of the servo loop.

Phasing by Paper Gap Pulse

Phasing has been described as the process of setting up the equipment so that the received copy is reasonably well centered on its blank. To accomplish phasing, a signal must be sent from one terminal to the other which can be used to phase the scanning motors. The system should be reliable, automatic, and require no transmission path other than that used to send the message portion of the signal.

Because the two transmitters operate alternately and are designed primarily for ease of loading, the copy may be placed in the drum without any regard for its position circumferentially and may have any phase relationship with respect to the drum. Such a situation requires that phasing precede every transmission.

The system used develops at the transmitter an 0.5-millisecond pulse from the

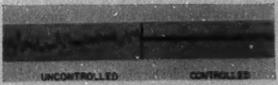


Figure 9. Comparison of the stability of an uncontrolled and a controlled motor

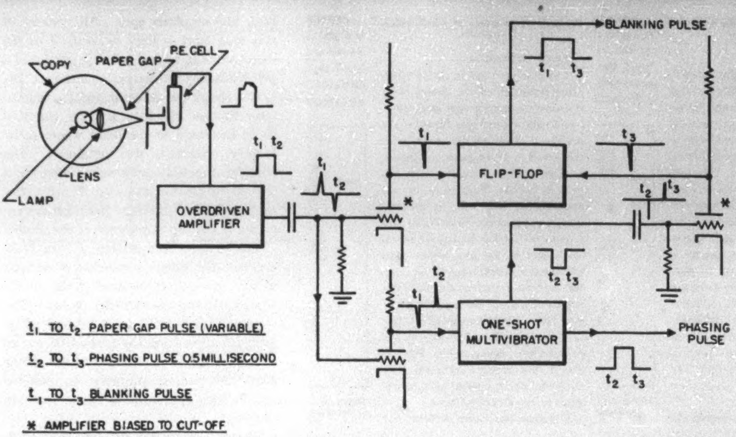


Figure 10. Transmitter pulse circuitry

paper gap, and with it modulates the facsimile carrier. The pulse-modulated carrier is transmitted automatically to the recorder for a period of 3 seconds just preceding each transmission. At the recorder, while the equipment is phasing, the motor is running at 1,830 rpm and a similar 0.5-millisecond pulse is developed. The transmitter pulse, after demodulation, and the recorder pulse are applied to a coincidence detector. Since the transmitter pulse occurs at a frequency of 30 pulses per second and the recorder at 30.5 pulses per second and they are each 0.5 millisecond wide, within a period of two seconds they will pass through coincidence. The pulse at coincidence will stop the phase-seeking device and change the recorder-motor speed from 1,830 to 1,800 rpm. The recorder now is running synchronously and in phase with the transmitter. The accuracy of phasing is dependent on the combined pulse widths. For two 0.5-millisecond pulses, 0.5 millisecond corresponding approximately to 1/8 inch on the copy, the theoretical maximum phasing variation is 1/4 inch. Wider phasing pulses would result in correspondingly larger phasing variations. It has been stated that the paper gap may vary from 1/8 inch up. It becomes obvious, therefore, that the paper-gap pulse itself should not be used as the phasing pulse, but rather a pulse of uniform width should be developed which has some definite relationship to an edge of the copy.

Figure 10 shows, in simplified form, the manner in which the phasing and blanking pulses are developed. An auxiliary photocell is installed on the outside of the scanning cylinder and a light source on the inside of the cylinder. Light is blocked off from the photocell by the copy except for that interval during which the paper gap passes by. Thus a pulse is generated for each drum revolution. This pulse may be of variable amplitude, depending on photocell sensitivity, light brilliance, and so forth, and is not entirely suitable for triggering purposes. To increase reliability, the pulse is fed to an overdriven amplifier where the grid is swung from zero bias to cutoff, thus obtaining a pulse of more nearly uniform amplitude and requiring only a minimum sensitivity from the photocell. The pulse is differentiated next and the resultant negative-going pulse, developed by the trailing edge of the gap, is caused to trigger off a 1-shot or mono-stable multivibrator. The multivibrator time constants are adjusted so that the resultant pulse, the phasing pulse, is of 0.5 millisecond duration.

Blanking Pulse Cuts Out Modulator

These circuits also develop the blanking pulse. The auxiliary photocell is so located that it sees the paper gap 0.25 millisecond before the scanning photocell. The differentiated pulse, resulting from the leading edge of the paper-gap pulse, is caused to trigger a flip-flop circuit and set up the beginning of the blanking pulse. The blanking pulse, in turn, shuts off the facsimile modulator 0.25 millisecond before the paper gap actually reaches the scanning cell.

The end of the blanking time is marked by a pulse differentiated from the trailing edge of the phasing pulse. This pulse restores the flip-flop circuit to its original condition and turns the modulator on again. The blanking pulse is, therefore, a summation in time of the paper-gap pulse and the phasing pulse. Because the phasing pulse is adjusted to be 0.5 millisecond long and anticipates the paper gap by 0.25 millisecond, the signal is blanked out for a period extending from 0.25 millisecond before until 0.25 millisecond after the paper gap has passed by the scanning photocell. This results in a loss of 1/16 inch at both right- and left-hand edges of the original copy.

Before describing the phasing process at the recorder it is necessary to explain some of the operations and their sequences in starting a recorder. The recorder is, in a sense, a complete slave of the transmitter, since all necessary operations are initiated and controlled from the transmitter; see Figure 11. A signal, which we shall call stand-by tone, is sent from the transmitter at all times the equipment is idle. On initiating a transmission, the stand-by tone is removed and during a period of 7 seconds the recording motor is started on commercial alternating current and transferred to run from the amplifier at 1,830 rpm. This speed, which we shall call the phase-seeking speed, is obtained by applying the standard 60 cycles to a rotatable transformer which is continuously rotated at the rate of 1 revolution per second by a small clock motor until the equipment is phased. In this way the standard 60 cycles is converted to 61 cycles before it is applied to the power amplifier and motor.

Imbedded in a flywheel on the motor shaft is a small piece of a permanent magnet which causes a pulse to be generated in a coil which is mounted adjacent to the flywheel. The physical relationship between the permanent magnet, coil, and styli is such that a stylus is always begin-

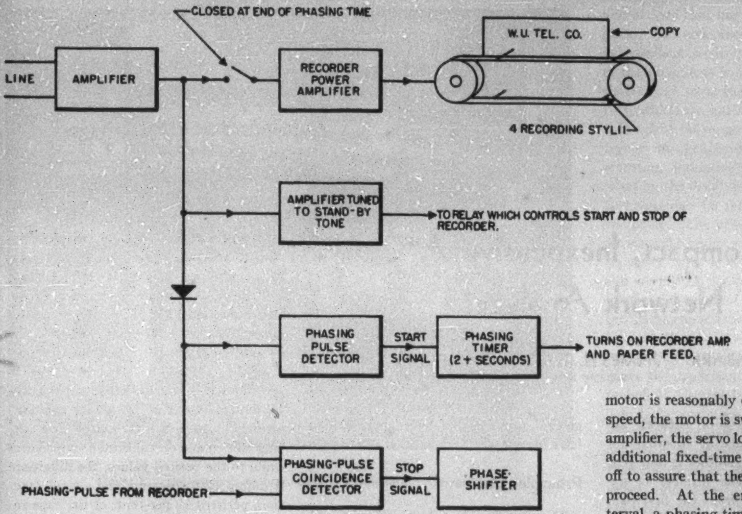

CLOSED AT END OF PHASING TIME
W.U. TEL. CO.
COPY
LINE
AMPLIFIER
RECORDER POWER AMPLIFIER
4 RECORDING STYLII
AMPLIFIER TUNED TO STAND-BY TONE
TO RELAY WHICH CONTROLS START AND STOP OF RECORDER.
PHASING PULSE DETECTOR
START SIGNAL
PHASING TIMER (2 + SECONDS)
TURNS ON RECORDER AMP. AND PAPER FEED.
PHASING-PULSE COINCIDENCE DETECTOR
STOP SIGNAL
PHASE-SHIFTER
PHASING-PULSE FROM RECORDER

Figure 11. Recorder scanning and signal detection

ning the scanning of the left margin of the copy while the pulse is being generated.

Phasing Interval Fixed by Timer

Upon removal of the stand-by tone, the phase-detector circuit becomes sensitive to phasing pulses. This circuit causes a relay to operate when phasing pulses begin arriving and initiates the phasing interval timer at the recorder. This timer, theoretically, needs to be set for no more than 2 seconds; however, to be positive that a minimum 2-second period is always measured off, it normally is set for about 2.5 seconds.

Some time during the phasing period the coincidence detector is triggered, indicating that the equipment has been phased, and it operates a relay removing the motive power from the clock motor which turns the rotatable transformer. At the end of the time measured off by the phasing interval timer, the recording amplifier and paper feed motor are turned on and, a fraction of a second later, the facsimile signals begin to appear and actual message recording commences.

If the transmitters are operated alternately, the stand-by tone is applied to the line between messages for only about 1/10 second and is immediately followed by phasing pulses. The 1/10-second tone not only initiates the cutoff and delivery of the message but the complete recycling of the recorder with the exception of the initial motor starting period.

To simplify the procedure required for the operation of a facsimile channel, a certain amount of control equipment is required. In the present instance, 13 relays and 3 timers are among the items required at the transmitting terminal. Their operation will not be described in detail but only insofar as they affect the apparent performance of the equipment.

12.5-Kc Tone for Stand-by

As has been stated previously, the transmitter applies a stand-by tone of 12.5 kc to the line during idle periods. The selection of 12.5 kc was based on a number of reasons, dependant somewhat on the 25-kc carrier and the nature of facsimile signals. Frequency choice was based on the premise that, with a 25-kc carrier, 12.5 kc was the frequency of least incidence at appreciable amplitude when considered either as a direct component or when modulated onto the carrier. This fact, in addition to the sending of this one frequency at full power, makes possible a positive selectivity on an amplitude basis against signals which may result from facsimile copy. The absence or presence of this tone on the line determines whether the recorder runs or stands idle.

The tone is removed by a relay which operates when a transmitter door is opened and closed. Closure of the door also starts the transmitter. After the voltage across the motor comes up to a predetermined value indicating that the motor is reasonably close to synchronous speed, the motor is switched to the power amplifier, the servo loop is applied, and an additional fixed-time interval is measured off to assure that the recorder is ready to proceed. At the end of this time interval, a phasing timer is started and the phasing pulse amplifier is turned on. Transmission of phasing pulses continues for 3 seconds, because at the recorder the phasing-pulse detector must operate, then a 2.5-second period is measured off for phasing, and, finally, the platen must bring the paper into contact with the styli. The time required is just a fraction under 3 seconds. At the end of the phasing time, scanning is initiated by turning on the facsimile amplifier and line-feed motor.

While the first transmitter is running, the second may be loaded and started. It will come up to speed and run idle until after the first message is completed when a switch is tripped causing the transmission of stand-by tone for about 1/10 second followed immediately by the transmission of phasing pulses from the second transmitter and in turn by the message. So long as the drums are kept loaded, only the 3 seconds of phasing time are lost between messages.

The transmitter and recorder are described in detail elsewhere.[1,2]

REFERENCES

1. A HIGH-SPEED TELEFAX RECORDER. D. M. Zabriskie. *Western Union Technical Review* (New York, N. Y.), volume 6, number 2, April 1952.
2. A HIGH-SPEED FACSIMILE TRANSMITTER, L. G. Pollard. *Western Union Technical Review* (New York, N. Y.), volume 6, number 2, April 1952.

No Discussion

A Compact, Inexpensive A-C Network Analyzer

EDWARD W. KIMBARK
FELLOW AIEE

JAMES H. STARR
FELLOW AIEE

JAMES E. VAN NESS
STUDENT MEMBER AIEE

THE great usefulness of network analyzers (a-c calculating boards) for making power-system studies is now generally recognized. The number of a-c boards has increased from the first two, which went into service in 1929.[1,2] to more than 25 at present.[3] Most of the existing a-c boards are in the United States, but there is at least one each in Canada, England, Brazil, and India. The studies made on these boards undoubtedly have resulted in economies in capital expenditures in power systems equal in value to many times the cost of the studies themselves.

An a-c board of average size (see Table I) and of any of the designs now most widely used would cost about $150,000 at present prices and would occupy a room having a floor area of about 700 square feet. In addition to this space, another room is needed for the motor generator set which must be connected by conduits to the board itself.

Although the existing types of a-c boards[1–6] have proven invaluable, a great need exists for a much more compact and less expensive board. To meet this need, the authors have designed a new a-c board, which, they estimate, can be built to sell commercially at a price of about 20 per cent of the price of a board of conventional design. It occupies a floor space of about 60 square feet in area; see Figure 2. It has no motor-generator; it can operate from any 50- or 60-cycle single-phase circuit and requires only 1,000 volt-amperes. It can be operated by one man. Such a board has been built at Northwestern University and is used there for instruction of students and for use by power companies in solving their network problems.

This board will be described after some general principles underlying its design have been stated.

Principles Used

The considerable reduction in size and cost was made possible by the following means:

1. The use of uncalibrated impedance units, which are set with the aid of instrument readings.
2. The representation of impedances by their conjugates.
3. The use of a low power level.
4. The use of standard mass-produced radio-electronic parts.

The first two of these features are covered by United States patents issued to one of the authors.[8,9]

The four items will be discussed in more detail.

UNCALIBRATED IMPEDANCES

Uncalibrated impedances, set by instrument, have been used for some time on d-c boards, effecting a great reduction in size, weight, and cost. The same principle has been applied to the new a-c board with equal success.

On the a-c board the impedance units consist of rheostats either in series or in parallel with adjustable reactors, and the units can be set to the desired values of either impedance or admittance in several alternative ways by the use of instruments. On the board described in this paper, the wattmeter-varmeter is used for this purpose. If an impedance is to be set, constant current of unit value (or multiple or submultiple thereof) is supplied to the impedance unit from a constant-current source, and the resistance and reactance are read on the wattmeter-varmeter with the watt-var switch in the "watt" and "var" positions, respectively. If an admittance is to be set, constant voltage of unit value is supplied to the impedance unit, and the conductance and susceptance are read on the meter in the "watt" and "var" positions, respectively.

Although it is true that setting uncalibrated impedance units with the aid of instruments takes a little longer than setting the dials of calibrated impedance units to the desired values, the difference is unimportant because only a small fraction, perhaps 10 per cent, of the time required for a study made on an a-c board is consumed in setting the impedance units.

An advantage of setting the impedance units by instrument, in addition to that of reduction of size and cost of the board, is that poor switch contacts and the like will be discovered while the unit is being set instead of later, during the course of a study, after hours have been spent in obtaining erroneous readings.

It is apparent that a considerable reduction in size and cost can be effected by using a continuously variable, uncalibrated rheostat plus several fixed resistors, the resistance of which need not be accurate, in place of two or three decades of resistors each step of which must have the correct resistance within a small tolerance. Likewise, a reduction can be effected in size and cost of inductive reactors by using uncalibrated adjustable

---

Paper 52-113, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1952. Manuscript submitted October 22, 1951; made available for printing December 18, 1951.

EDWARD W. KIMBARK is at the Instituto Tecnológico de Aeronáutica, São José dos Campos, São Paulo, Brasil, JAMES H. STARR is in La Grange, Ill., and JAMES E. VAN NESS is at Northwestern University, Evanston, Ill. Mr. Kimbark was connected with Northwestern University when this paper was written.

The authors wish to acknowledge the important contributions of Evan Richard Owen, formerly Instructor, Department of Electrical Engineering, Northwestern University, now with the General Electric Company, Schenectady, N. Y. His design of the instrument amplifiers was vital to the success of the project.

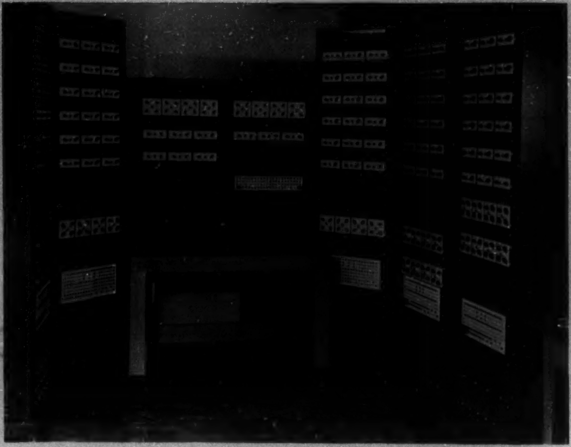


Figure 1. The network analyzer built at Northwestern University

reactors. However, a still greater reduction can be made by the use of conjugate impedances.

Conjugate Impedances

One of the most difficult and important problems in the design of a conventional a-c board has been that of designing the adjustable inductive reactors of the line units. These reactors must have at each setting a reactance which is substantially independent of current and must also have low and preferably constant ratio of resistance to reactance, that is, high Q. These requirements have been met, in the usual types of a-c board, by the use of cores of special alloy steel with air gaps, of a larger size of wire than required for current-carrying capacity, and of a frequency between 400 and 500 cycles. The reactors of the load units do not need to meet requirements as stringent as those for the reactors of the line units, yet both kinds of reactors are constructed similarly. Because of the large number of line and load units on an a-c board, the size, weight, and cost of the reactors are important in fixing the size, weight, and cost of the entire board. The size of the reactors can not be decreased beyond a certain point without making their ratio of resistance to reactance too high.

Two methods of obtaining good reactors at lower cost have already been put into practice. The first method is to use a higher frequency (10,000 cycles has been used).[6] The second method is to use Professor Phillips' phase-shift principle at 60 cycles.[7] The authors believe, however, that the use of conjugate impedances is preferable to both of those methods from the standpoints of size and cost.

Our new a-c board represents inductive reactance by capacitive reactance. Each line or load unit has resistors and capacitors. The reactance of the capacitors is independent of current, and the resistance of the capacitors is negligible. These resistors and capacitors are of the radio type, which are small and inexpensive.

It might be supposed that the advantage of using capacitive reactance for representing inductive reactance would be offset by the disadvantage of using inductive reactance for representing capacitive reactance. The offsetting is only partial for several reasons. First, the number of capacitor units on an a-c board of average size is only 24 per cent of the combined numbers of line and load units. Second, the shunt capacitance of a power system has less effect on the operation of the system than does the series inductance, and hence the capacitance need not be represented as accurately as the inductance. Third, the variation of voltage across the shunt capacitance during the normal operation of a power system is much less than the variation of current through the series inductance; therefore, the shunt inductors used to represent capacitance under the conjugate-impedance principle need not be designed for as small a variation of reactance with current as is required for the line reactors on a conventional board. Fourth, most of the shunt capacitance is assumed to be connected to busses where there are also shunt loads, and the losses in the shunt inductor which represents the capacitance can be considered as part of the active power of the load. In view of these facts, there is still an immense advantage in using conjugate impedances on an a-c board.

Low Power Level

The power level of the first a-c boards was determined by the burden of the available measuring instruments. This limitation was later removed by the employment of negative-feedback vacuum-tube amplifiers between the network and the instruments.[10,11] The power level was then probably determined on some boards by previous practice and on others by the

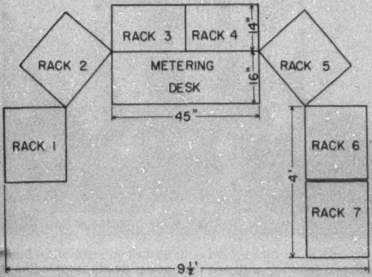


Figure 2. Dimensioned floor plan of the network analyzer shown in Figure 1

Table I

Number of Circuit Units	Type
12	Generators
99	Lines
40	Loads
30	Shunt Units*
18	Autotransformers
15	Mutual transformers
214	

* Representing capacitive reactance by inductive reactance.

requirement for low-loss reactors. The use of conjugate impedances permits a further decrease in the power level to a point governed by the difficulties of shielding the circuits against undesired pickup and of designing instrument amplifiers of sufficiently constant gain yet with simple circuits and with freedom from oscillation.

The base power used on the new board is only 0.5 watt, which is lower than that of any previous a-c board. The use of low base power permits the use of small, inexpensive resistors in the line and load units. It likewise reduces the required rating of the generator units and makes an electronic type of generator unit both feasible and attractive. The use of a high base impedance (toward which the designer is impelled in order to decrease the capacitances required by the conjugate-impedance principle) together with a low base power gives a low base current, which permits the use of small switches, plugs, and jacks with freedom from troubles caused by high contact resistances.

USE OF MASS-PRODUCED PARTS

The conjugate-impedance principle and the low base power permit the line and load units to be built from ordinary radio resistors, rheostats, and capacitors. Ordinary jacks, plugs, switches, panels, cabinets, and other parts have been used throughout the board wherever possible. Such parts are not only inexpensive but are readily available from several different manufacturers. Few components were designed especially for this board.

Description of the New A-C Network Analyzer

BASE QUANTITIES

The following base quantities are used on the board:

1. Voltage—50 volts.
2. Current—10 milliamperes.
3. Impedance—5,000 ohms.
4. Power—0.5 watt.
5. Frequency—120 cycles.

The frequency of 120 cycles is obtained from the 60-cycle supply by means of a frequency doubler. It was chosen because it is the highest multiple of 60 cycles at which standard measuring instruments are guaranteed to have their rated accuracy. The base impedance of 5,000 ohms resulted from two decisions: that the least reactance setting of a line unit should be 2 per cent of base and that for the desired compactness the total capacitance of a line unit should not exceed about 13 microfarads.

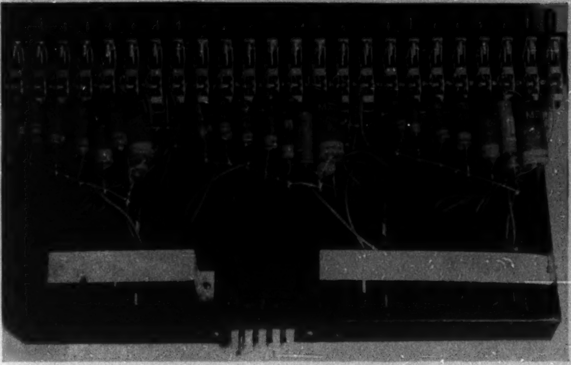


Figure 3. Top view of a panel with three line units mounted on it

NUMBER OF UNITS

The number of circuit units of each type, shown in Table I, was decided upon after a tabulation had been made of the numbers of units on nearly all the existing a-c boards.[3,12,13] The number was taken approximately equal to the average number of units of corresponding type on the existing boards.

LINE UNITS

The line units have capacitive reactance, representing inductive reactance, and resistance which is normally in series with the reactance. However, a 9-position rotary switch makes the following connections possible: (1) open circuit, (2) reactance only, (3 to 9) resistance and reactance in series, with the resistance increasing in fixed steps. By means of these steps and a stepless rheostat, the resistance is continuously adjustable from zero to 112 per cent of base. The capacitance is adjustable in 0.001-microfarad steps from zero to 13.332 microfarads, nominal values. The minimum reactance, corresponding to maximum capacitance, is about 2 per cent of base. The maximum usable reactance depends upon the accuracy with which the reactance must be set. At a setting of 130 per cent of base impedance the smallest step of adjustment is 0.5 per cent of the setting; at 260 per cent of base, 1 per cent; 520 per cent, 2 per cent; 1,300 per cent, 5 per cent; and so on.

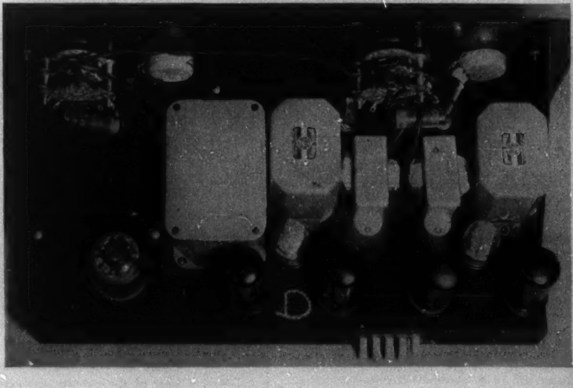


Figure 4. Top view of a panel with two generator units and their power supply mounted on it

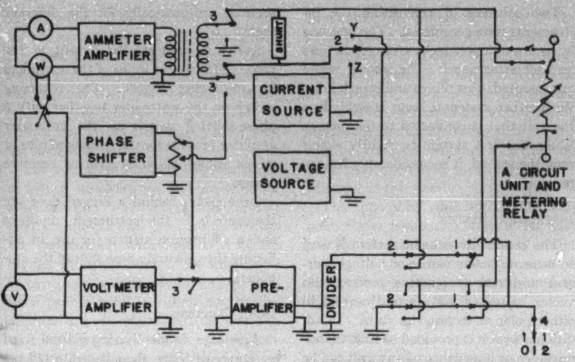


Figure 5. Simplified diagram of the instrument switching system. Switch 1 is the voltage selector switch, Switch 2 connects the system for measuring impedance or admittance, Switch 3 connects the instrument phase shifter, and Switch 4 selects the sequence bus for voltage measurements

Much study was given to various arrangements of switches and capacitors which would cover the desired range of capacitance with no gaps while utilizing capacitors of commercial nominal capacitance and tolerance. For uncalibrated units, a decade arrangement is not necessary. Four-position lever switches were found to occupy less panel area than any arrangement of rotary switches that would accomplish the desired results. Each switch provides capacitances of relative values 0, 1, 2, and 3, the 1 and 2 being obtained by separate capacitors and the 3 by these in parallel. On successive switches the unit value is, in turn, 0.001 microfarad, 0.003 microfarad, 0.01 microfarad, 0.03 microfarad, 0.1 microfarad, 0.3 microfarad, 1 microfarad, and 3 microfarads.

Now consider the problem of bridging the gap between the steps of 0.1-microfarad nominal value. The smaller capacitors on the next two switches provide a total nominal capacitance of 0.12 microfarad. Even if all these capacitors should be 5 per cent low, giving only 0.114 microfarad, they would be just adequate to bridge the gap which would occur if the 0.2-microfarad capacitor were 5 per cent high and the 0.1-microfarad capacitor were 5 per cent low. The other nominal 0.1-microfarad steps cannot exceed 0.105 microfarad, and the nominal 0.3-microfarad steps can be bridged more easily than the 0.1-microfarad steps. Thus capacitors of ±5 per cent tolerance are adequate for the worst condition.

All capacitors bought for the line and load units were measured on a bridge and sorted into groups according to the following ranges of error: −10 to −5 per cent, −5 per cent to 0, 0 to +5 per cent, and +5 to +10 per cent. All capacitors for the same unit were taken from the same group as far as possible; otherwise, from adjacent groups.

Figure 3 shows three line units which are mounted on a panel $5^{1}/_{4}$ inches by 19 inches. Their rating, determined by the resistors, is 6.3 per-unit current.

LOAD UNITS

The load units have capacitive reactance, representing inductive reactance, and resistance which is normally in parallel with the reactance. An 8-position switch makes possible the following arrangements: (1) open circuit, (2) reactance only, (3, 4, and 5) different ranges of resistance in parallel with reactance, and (6, 7, and 8) different ranges of resistance in series with reactance. Resistance only can be obtained on any of the parallel connections by setting the capacitance at zero. With the parallel connections, the conductance is adjustable continuously from 0.9 per cent to 200 per cent of base, and the susceptance from 0.4 per cent to 125 per cent of base in steps of 0.4 per cent of base.

Four load units are mounted on a panel 7 by 19 inches. These units are rated at 200 per cent of base voltage.

SHUNT UNITS

The units used to represent the shunt capacitance of the power system are called shunt units, because to call them capacitor units would be confusing in view of the employment of the conjugate-impedance principle. The shunt units must provide adjustable inductive susceptance. Each shunt unit consists of a center-tapped inductive reactor in parallel with a bank of capacitors and switches similar to those used in the load units. The inductive reactor has, for the full winding, a susceptance of 10 per cent of base and a Q of 15, giving a conductance of 0.7 per cent of base. On half the winding it has a susceptance of 40 per cent and a conductance of 4 per cent. Lower values of inductive susceptance are obtained by shunting the reactor with capacitors. The Q of the parallel combination is lower than that of the reactor alone, as the conductance remains constant. The smallest increment of susceptance is the same as on a load unit. The susceptance of the reactor is constant within ±1.2 per cent for a variation of voltage from 50 per cent to 200 per cent of base. The reactor has a special alloy core and weighs 8 pounds. Six shunt units are mounted on a panel 19 by $5^{1}/_{4}$ inches.

AUTOTRANSFORMERS

The autotransformers provide a buck or boost adjustable by dial switches from 0 to 25 per cent by 0.5-per-cent steps. Rated secondary current is 630 per cent of base (equal to rated current of the line units). The open-circuit admittance of the primary winding, resonated with a shunt capacitor, is $1.4+j0.0$ per cent of base. The equivalent series impedance, on the 100-per-cent-voltage side, at 25-per-cent boost is $2.2+j0.0$ per cent of base. Each autotransformer weighs $2^{1}/_{4}$ pounds.

MUTUAL TRANSFORMERS

Two-winding transformers of ratio 1 to 1 are provided for use in various equivalent circuits, particularly that for two coupled transmission lines not bussed at either end. The open-circuit impedance of the transformer at 40 per cent of base voltage is $1{,}380+j880$ per cent of base. The short-circuit impedance is $4.3+j0.7$ per cent of base. The interwinding capacitive susceptance is 0.1 per cent of base. The rated current is 650 per cent of base. The transformers weigh $2^{1}/_{4}$ pounds each.

When the transformers are used together with line-impedance units for representing the self- and mutual-impedances of coupled lines, it is possible to set the desired impedances accurately by instrument in spite of the imperfections of the transformer. Thus, when mutual impedance is to be adjusted, the wattmeter is connected to receive the current of one winding of the transformer and the voltage of the other winding in shunt with a line unit. Then the line unit is adjusted until the desired value of mutual impedance is indicated. The self-impedance is measured as the impedance of a line unit in series with one winding of the shunted transformer.

GENERATOR UNITS

Each generator acts as a source of single-phase, 120-cycle voltage, adjustable independently in phase and in magnitude but substantially independent of the output current.

All generators are supplied with 120-cycle 2-phase voltage from the frequency doubler voltage regulator. The phase adjustment of each generator consists of four 90-degree steps, obtained by switching from one phase of the supply line to the other and by reversal of polarity, and a continuously variable shift of 150 degrees. The output of the phase-shift circuit goes to a voltage-divider rheostat which provides stepless adjustment of the magnitude to any value from zero to two times base voltage. Next comes a 2-stage power amplifier having a large amount of negative feedback. The output impedance of the generators is less than 0.5 per cent. The harmonic content of the output voltage is less than 4 per cent, this being almost entirely made up of second and third harmonics. Two generators and the d-c power supply for their tubes are mounted on a panel $8^3/_4$ by 19 inches. One unit is shown in Figure 4.

A modification of the feedback circuit permits the generator to function as a constant-current source instead of as a constant-voltage source. Only one of the generators is thus equipped. It is used as a constant-current source for the measurement of impedances.

SCHEME OF CONNECTIONS

Any one of several schemes of connections could be used. With the reduced size of this board it would be possible to make interconnections by means of jacks placed on the outside edge of the unit panels. Then the meters could be connected through a 3-pole jack on the panel. This scheme has the merit of requiring no other connections to the unit than those on the front panel.

On the board built at Northwestern University the leads of each unit are brought out through connectors on the back of the units to a central connection panel on each rack. A system using banana-type plugs and jacks is used for interconnection on this panel. Adjacent racks have their connection panels at the same height so that connections between racks can be made either directly from one to the next or through busses that are common to each rack.

Two schemes of connecting the instruments were considered. The first was by means of 3-pole jacks arranged in rows on a metering panel. The second, which was adopted, uses relays and punch keys. With either system a lamp is lighted on the unit that is connected to the meters. While the first system is slightly slower than the second, it is considerably less expensive.

INSTRUMENT SYSTEM

The central instrument system is used to measure scalar values of voltage, current and real or reactive power; also vector values of voltage and current in either polar or rectangular form. In addition a switch is provided so that the admittance or impedance of any unit can be measured directly without removing that unit from the network.

The instrument system has three indicating instruments, a voltmeter, ammeter, and wattmeter-varmeter. They are driven by two amplifiers, one for current and one for voltage. A phase-shift circuit is switched into the voltmeter amplifier when it is desired to measure vars.

A synchro having nonsalient poles and high-impedance 3-phase windings on both stator and rotor is used as an instrument phase shifter. A switch is provided for connecting its output to the input of either the voltage or current amplifier. It is driven from a single-phase source through a phase-splitting network.

INSTRUMENT SWITCHES

Figure 5 gives a simplified diagram of the instrument switching system. Switch 1 provides for connecting the instruments to read the power into a unit, the power out of a unit, and the power lost in the unit. This is accomplished by switching the voltmeter leads to read the voltage at the input terminals, output terminals and the voltage drop across the unit.

Switch 2 connects the instruments for measuring admittance or impedance directly with the instruments; at the same time, it disconnects one end of the unit being measured from the network, thus making it unnecessary for the operator to disconnect the leads to that unit's terminals. In addition to the sections shown, other sections, in conjunction with an impedance or admittance range switch, choose the proper values of the current shunt and the proper ranges for the current and voltage source. The current source and the voltage source are actually the same unit and are also used as a generator during a network study. Switch 2, acting through relays, changes their connections automatically for the different uses.

Switch 3 connects the output of the phase shifter to the input of the amplifiers for measuring angles. The reversing switch on the wattmeter together with a phase-shifting switch in the voltmeter amplifier provide for the reading of plus or minus values of either real or reactive power.

On a panel behind a hinged door are the controls for the instrument amplifiers and a calibration switch for use in adjusting the gain and phase shift of the amplifiers.

INPUT CIRCUITS

A voltage divider having a total fixed resistance of 5 megohms is connected between the points whose potential difference is to be measured. A 0.5-megohm portion at one end of the divider is tapped, and the taps go to the voltage range selector switch on the output side of the divider. Ranges of 0.075, 0.15, 0.3, 0.75, 1.5, or 3 per-unit voltage are provided.

A tapped shunt having a different value of resistance for each current range is inserted in series with the circuit unit whose current is to be measured. In order to prevent errors from switch-contact resistance, the taps go to two separate wafers of the current range selector switch, one wafer on the input (main current) side of the shunt, the other on the output side. On each range, the voltage across the shunt for full-scale deflections of the ammeter is 37.5 millivolt (0.075 per cent of base voltage). Ranges of 0.1, 0.5, 1, 2, 5, or 10 per-unit current are provided. The voltage across the shunt is fed into the current amplifier through an electrostatically shielded input transformer since the shunt will usually be at a different voltage level than the amplifier chassis.

Although the voltage divider has a shunt conductance of only 0.1 per cent of base, the total shunt admittance introduced with the insertion of the instruments in the circuit to be measured is greater than this. The total admittance varies somewhat with the position of the voltage selector switch but in no case exceeds 0.20 per cent of base. The additional admittance arises from the stray capacitance of the metering circuit. Most of this is resonated out, but such a large value of inductance is needed that the Q of the circuit is low enough to introduce appreciable conductance.

AMPLIFIERS

The ammeter amplifier has three stages of voltage amplification and one of power amplification. There are two minor feedback loops and one over-all feedback loop to insure linearity. Between the transformer across the shunt and the input to the amplifier there is a two-time-constant resistance-capacitance low-pass filter to reduce any errors that might arise from harmonic currents.

The voltmeter amplifier has four stages of voltage amplification and one of power amplification. The voltage is fed directly from the divider taps to a pentode to keep the capacitance across the taps of the divider at a minimum. The amplifier is in two sections, each having its own feedback path around it. Between the two sections are two phase-shift circuits, one of +45 degrees and one of −45 degrees, which are switched in by the watt-var switch.

Conclusion

Compact, inexpensive a-c network analyzers such as the one described are capable of handling all the problems arising in power system operation that can be handled by other types of boards. While a slight sacrifice in operating time has been made, this is more than offset by the advantages of economy and reliability of operation. The design is quite flexible in that many features, such as instruments on generators, the means of interconnection, and the method of connecting the instruments into the circuit to be measured, can be changed without affecting the basic principles used to achieve the advantages of this board.

This type of board should enable many more power companies to own one than now do. The increased number of studies thereby made possible should result in further increase in the economy and reliability of electric power supply.

References

1. The M.I.T. Network Analyzer: Design and Application to Power System Problems, **H. L. Hazen, O. R. Schurig, M. F. Gardner.** *AIEE Transactions*, volume 49, July 1930, pages 1102–14.

2. An Alternating-Current Calculating Board, **H. A. Travers, W. W. Parker.** *Electric Journal* (East Pittsburgh, Pa.), volume 27, May 1930, pages 266–70.

3. An Improved A-C Network Analyzer, **W. A. Morgan, F. S. Rothe, J. J. Winsness.** *AIEE Transactions*, volume 68, part II, 1949, pages 891–97.

4. A New A-C Network Analyzer, **H. P. Kuehni, R. G. Lorraine.** *Electrical Engineering* (*AIEE Transactions*), volume 57, February 1938 pages 67–73. Discussion, *Electrical Engineering* (*AIEE Transactions*), volume 57, July 1938, pages 418–22.

5. The Modern A-C Network Calculator, **W. W. Parker.** *Electrical Engineering* (*AIEE Transactions*), volume 60, November 1941, pages 977–82. Discussion, pages 1395–98.

6. A New Design for the A-C Network Analyzer, **J. D. Ryder, W. B. Boast.** *Electrical Engineering* (*AIEE Transactions*), volume 65, October 1946, pages 674–80. Discussion, pages 1162–65.

7. Impedance-angle Shift, **Emory B. Phillips.** United States patent number 2,487,942, November 15, 1949.

8. Uncalibrated Impedances, **James H. Starr.** United States patent number 2,301,470, November 10, 1942.

9. Conjugate Impedances, **James H. Starr.** United States patent number 2,523,453, September 26, 1950.

10. A Stabilized Amplifier for Measurement Purposes, **H. A. Thompson.** *Electrical Engineering* (*AIEE Transactions*), volume 57, July 1938, pages 379–84.

11. An Amplifier-Wattmeter Combination for the Accurate Measurement of Watts and Vars, **G. S. Brown, E. F. Cahoon.** *Electrical Engineering* (*AIEE Transactions*), volume 58, November 1939, pages 593–98.

12. Today's Network Calculators Will Plan Tomorrow's System, **Dan Braymer.** *Electrical World* (New York, N. Y.), volume 125, January 5, 1946, pages 52–54.

13. Use of A-C Network Calculator in Planning and Operating Electric Power Systems, **P. O. Bobo, H. A. Travers, E. E. George.** *Report* 328, International Conference on Large High Voltage Electric Systems, 12th Session, Paris, France, 1948.

Discussion

Lester B. Le Vesconte (Sargent and Lundy Engineers, Chicago, Ill.): This development appears to be very practical. The greatest deficiency in the present installation is the lack of instruments on the units. The paper indicates that instruments could be added to the generators but should be added also to the loads. Each instrument requires an amplifier which may seem to be prohibitive in cost. What seems to be needed is an inexpensive amplifier recognizing that accuracy is not important on these unit instruments. The only requirements are consistent indications and a low temperature error to save time when first energized.

R. B. Squires (Westinghouse Electric Corporation, East Pittsburgh, Pa.): The paper shows several ingenuous devices for reducing the size and first cost of a network calculator to a minimum. The operating cost must also be considered, not only for equipment replacement but for engineering time per study. Mass produced radio parts are designed for uses in which a long trouble-free life and constant low contact resistance are not of the great importance they are in calculators.

The use of uncalibrated impedances increases the time required to start a study. In addition, the omission of most of the refinements of modern calculators—load adjusters, generator instruments, load voltmeters, and plotting table—will increase the operating time for a study using 12 generators by several hundred per cent. During a critical shortage of engineers, this loss of engineering time must be balanced against the higher first cost of a high-productivity design, or the nominal rental charges of such existing calculators.

However, in its use in the university, the unit described in the paper does have an important place because of its use in teaching students the theory of circuits and power system analysis, and as an instrument for postgraduate research.

J. M. Hirst (Thayer School of Engineering, Hanover, N. H.): The calculator design described in this paper is an impressive step in the direction of reduced cost and simplicity of construction. These factors are of particular importance where the use factor of a calculator may be relatively low. The authors, in their conclusion, have pointed out the possibility of the smaller power company owning its own board. Another possibility, and one in which the use factor would probably be considerably lower, is for educational institutions where participation by utilities is not feasible.

From the educational standpoint a calculator can serve at several stages of student development. It can, especially if of simple construction, function as part of the laboratory for a-c circuits work. In conjunction with courses in power systems it permits students to check their own network computations and to determine directly the effects of varying parameters or operating conditions. It may also be useful for research, or for graduate students' thesis projects.

These considerations raise the interesting question as to the possibility of further substantial reduction in cost. One wonders if the relative cost of the instrumentation is low enough so that an over-all reduction in the number of units would result in anywhere near a proportional reduction in cost. It also would seem that for educational use the substitution of a jack-type metering panel for the relay system, as suggested in the paper, would be a justifiable economy.

The use of conjugate impedances appears to be a very satisfactory way of avoiding the problem of a low frequency, high Q, linear inductor. The problem still remains, however, associated with the inductor units which represent capacitance; and one might wonder if the Q values of 10 to 15 mentioned in the paper are high enough to provide good accuracy when these units are used as the shunt branches of the pi equivalent of a long line, where the admittance angle should be close to 90 degrees. In such a case it would not appear appropriate to consider the extra loss of the shunt unit as part of the power load, particularly in the case of the sending end branch.

However, if these extra losses are treated as errors, without adjustment of the load, the results will probably still be within very satisfactory limits of accuracy. An investigation of the effect, on a typical receiving circle diagram for a 266-mile line, of an angle of 85 degrees (Q between 10 and 15) bears this out. Considering all parameters accurate except the admittance angle of the shunt branches, the circular locus was found to be displaced from about 0.25 per cent at low power angle to about 1.25 per cent near the steady state power limit. This displacement would, in general, result in larger per cent errors in actual complex power quantities, since the usual circles are considerably closer to the P-Q origin than to their own center. At normal values of transmitted power the volt-ampere error would be of the order of 1 per cent of the power and the error in power angle not more than 1.5 degrees.

Edward W. Kimbark, James H. Starr, James E. Van Ness: The network analyzer described in this paper was designed primarily for educational use where some sacrifices in speed and convenience of operation in favor of economy were highly desirable. The omission of meters on the generator and load units was one such change from standard network analyzer practice. While this does increase the operating time slightly, it does not increase it by a factor of several hundred per cent. Unlike most network analyzers, it is possible to adjust the generator and load units while simultaneously reading the main set of instruments due to the proximity of all the units to the main metering desk. Therefore, the only increase in time is that required to transfer the main instruments from one unit to another.

With a careful choice of the parts used, it was felt that mass produced electronic parts would give satisfactory service in the analyzer. Their use was felt even more justified because of the uncalibrated impedance principle which would immediately detect any failure of parts in the units.

It is hard to accurately separate the total cost of the instrument system from that of the total board. The cost of the parts for the instrument system described in this paper, not including the connection system whose cost would vary with the number of units, was less than 10 per cent of the cost of the parts for the total board. However, a large part of the development time for the board was spent on the instrument system.

Synthesis of the Transfer Function of 2-Terminal Pair Networks

ROBERT KAHAL
MEMBER AIEE

Synopsis: The open-circuit transfer function $\theta(p)$ of the 2-terminal pair network, defined as the ratio of the output voltage to the input voltage, must be regular in the right-half p plane. Poles of $\theta(p)$ on the imaginary p axis, if any, must be simple with residues having zero real part. These conditions are sufficient to realize $\theta(p)$ to within a constant factor by a symmetrical lattice network. If, in addition, $\theta(p)$ has no zeros in the right-half plane, then θ may be realized to within a constant factor by a ladder network. If $\theta(p)$ is further restricted by the condition $0<\theta<1$ whenever θ is real in the right-half plane, then θ may be realized completely by an L network. An algorithm for factoring the product of two positive real functions is given also.

THE open-circuit transfer function of the 2-terminal pair network is of some interest theoretically as well as in certain applications. Hence it is important to establish the form and properties which a physically realizable transfer function, as a function of frequency, may assume. This will be done here for certain classes of networks with suitable restrictions. The networks under consideration are assumed to be passive and constructed with a finite number of lumped, linear, and positive elements, R, L, and C.

General Considerations

Let the network be represented, as in Figure 1, by a box with two pairs of accessible terminals and with positive directions of voltage and current as shown. A mesh analysis, with mesh 1 taken through the left-hand pair of terminals and mesh 2 through the right-hand pair yields, the set of equations

$$I_1=E_1\frac{\Delta_{11}}{\Delta}-E_2\frac{\Delta_{21}}{\Delta} \tag{1}$$

$$I_2=E_1\frac{\Delta_{12}}{\Delta}-E_2\frac{\Delta_{22}}{\Delta} \tag{2}$$

where Δ is the determinant of the mesh equations and Δ_{ij} represents the indicated minor of Δ. For a passive network, $\Delta_{12}=\Delta_{21}$, hence from equations 1 and 2 it is clear that the network is completely specified by the three parameters given here:

Δ/Δ_{11}, the driving point impedance at terminals 1.
Δ/Δ_{22}, the driving point impedance at terminals 2.
$\Delta/\Delta_{12}=\Delta/\Delta_{21}$, the transfer impedance between terminals 1 and 2.

As Gewertz[1] has shown, all three functions cannot be chosen independently even though individually each may be an admissible network function.

When the right-hand pair of terminals is open circuit, $I_2=0$ and from equation 2 there results

$$E_2/E_1=\Delta_{12}/\Delta_{22} \tag{3}$$

The ratio $E_2/E_1=\theta(p)$ defines the transfer function of the network and is regarded as a function of the complex frequency variable $p=\sigma+i\omega$. Clearly $\theta(p)$ by no means completely specifies the network. For any given realizable function, $\theta(p)$, there exists a class of networks with the given transfer function but differing in other characteristics.

From the form $\theta(p)=\Delta_{12}/\Delta_{22}$ it is possible at once to make certain statements regarding θ. Thus, $\theta(p)$ is a rational function of p with real coefficients and so $\theta(p)$ is real when p is real. This condition is implied for all network functions encountered in this paper.

Since the zeros of Δ_{22} cannot have positive real parts $\theta(p)$ has no poles in $Re[p]>0$ (that is, in the right-half p plane). Moreover, since zeros of Δ_{22} on the imaginary axis must be simple, poles of $\theta(p)$ on $Re[p]=0$ (that is, on the imaginary p axis), are simple where $Re[\ ]$ denotes "the real part of."

There are, of course, other properties of the transfer function, but these will be adduced as required in connection with networks of specific form. For the sake of brevity the term "boundary" will be used to designate the imaginary axis in the complex p plane.

Positive Real Functions

From time to time it will be necessary to invoke certain properties of the driving point impedance, $Z(p)$. As is well known, $Z(p)$ must be a positive real function. Brune[2] has shown that the necessary and sufficient conditions for a rational function, $Z(p)$, to be a positive real function are

$Z(p)$ is real when p is real.
$Z(p)$ has no poles in $Re[p]>0$.
Poles on the boundary are simple, with positive, real residues.
$Re[Z(p)]\geqq 0$ on $Re[p]=0$.

Foster[3] has shown that the driving point impedance, $X(p)$, of a network containing inductances and capacitances only can be represented by one of the two expressions

$$X(p)=Hp\,\frac{(p^2-p_2^2)(p^2-p_4^2)\ldots(p^2-p_k^2)}{(p^2-p_1^2)(p^2-p_3^2)\ldots(p^2-p_{k-1}^2)}$$

$$X(p)=H\frac{1}{p}\,\frac{(p^2-p_1^2)(p^2-p_3^2)\ldots(p^2-p_{k-1}^2)}{(p^2-p_2^2)(p^2-p_4^2)\ldots(p^2-p_k^2)}$$

where H is a positive constant and p_j^2 is a negative real constant with $|p_{j+1}^2|>|p_j^2|$. Such a function will be called a "positive real reactance function."

Paper 52-146, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE South West District Meeting, St. Louis, Mo., April 15–17, 1952. Manuscript submitted December 28, 1951; made available for printing February 27, 1952.

ROBERT KAHAL is with Washington University, St. Louis, Mo.

This paper contains the principle results of a doctoral dissertation submitted to the Graduate Faculty of the Polytechnic Institute of Brooklyn (N. Y.). The author wishes to express his appreciation to Prof. R. M. Foster for his interest in this investigation and for suggesting the problem.

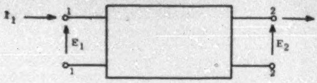


Figure 1 (left). Two-terminal pair networks

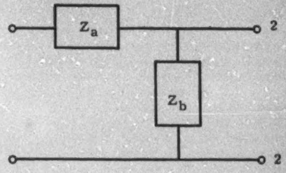


Figure 2 (right). L network

Transfer Function of the L Network

Let $\theta_L(p)$ denote the transfer function of the L network in Figure 2, where $Z_a(p)$ and $Z_b(p)$ represent arbitrary driving point impedances or positive real functions. It is understood that there is no mutual coupling between the series and shunt arms. The necessary and sufficient conditions which θ_L must satisfy can be obtained easily. A method for synthesizing the network will also be given.

Obviously, for the L network

$$\theta_L(p)=\Delta_{12}/\Delta_{22}=Z_b/(Z_a+Z_b) \tag{4}$$

and from the preceding discussion

$\theta_L(p)$ has no poles in $Re[p]>0$. Poles on the boundary must be simple and have residues with zero real part. (I)

The associated function, $Z_a\theta_L=Z_aZ_b/(Z_a+Z_b)$, is a positive real function since it represents a driving point impedance. Hence $Z_a\theta_L$ has no zeros in $Re[p]>0$ and zeros on the boundary must be simple. Hence, if θ_L has a zero in $Re[p]>0$, Z_a must have a pole of the same order in $Re[p]>0$. This is impossible since Z_a is a positive real function. If θ_L has a zero on the boundary of order $k(k>2)$, then Z_a must have a pole at that point of order $(k-1)$. This is likewise impossible for the same reason. Hence

θ_L has no zeros in $Re[p]>0$. Zeros on the boundary are at most of order 2. (II)

The reciprocal function is $1/\theta_L=1+Z_a/Z_b$. It is easily established that the product (or quotient) of two positive real functions cannot be real and negative in $Re[p]>0$. Therefore $1<1/\theta_L$ in $Re[p]>0$, and it follows that

$0<\theta_L<1$ in $Re[p]>0$. (III) This notation is used throughout to denote that condition III is satisfied when $\theta(p)$ is real in $Re[p]>0$.

From this it is clear that $0\leqq\theta_L\leqq 1$ at the origin and at infinity.

As a consequence of condition III, it is easy to show that $|\arg\theta_L|\leqq\pi$ on $Re[p]=0$. For, suppose θ_L is real and negative at some point p_0 on the boundary. Expanding θ_L in a Taylor's series about p_0 yields the expression $\theta_L(p)=\theta_L(p_0)+\theta_L'(p_0)(p-p_0)+\theta_L''(p_0)(p-p_0)^2/2!+\ldots$

Letting $\theta_L(p_0)=-a$, $\theta_L'(p_0)=k\epsilon^{i\vartheta}$, and $(p-p_0)=\rho\epsilon^{i\varphi}$, then for points sufficiently close to p_0, $\theta_L(p)$ is given approximately by $\theta_L\approx -a+k\rho\epsilon^{i(\vartheta+\varphi)}$ and the imaginary part of θ_L is approximately $Im\,[\theta_L]\approx k\rho\,\sin(\vartheta+\varphi)$. Since condition III requires that θ_L shall not be negative real in $Re[p]>0$, then $\sin\,(\vartheta+\varphi)\neq 0$ for $-\pi/2<\varphi<\pi/2$. Thus $\vartheta=\pm\pi/2$ and $Im[\theta_L]$ cannot change sign in passing through p_0 along the boundary. Hence the map of the p boundary cannot cross the negative real θ_L axis. In a similar fashion it can be established that the second statement in each of conditions I and II is a consequence of III for functions which, with their reciprocals, are regular in $Re[p]>0$.

Using only the properties I, II, and III of θ_L it is simple to show that the function $(1-\theta_L)$ also possesses these same properties. For, since the poles of $(1-\theta_L)$ are the same as the poles of θ_L, condition I is satisfied. When $0<\theta_L<1$, obviously we also have $0<(1-\theta_L)<1$, hence condition III is satisfied. Since θ_L cannot be unity in $Re[p]>0$, the function $(1-\theta_L)$ can have no zeros in $Re[p]>0$. Hence $(1-\theta_L)$ also satisfies condition II. This follows from the last statement of the preceding paragraph.

The properties of $(1-\theta_L)$ were derived in this fashion rather than by noting that it is the transfer function of the network in which Z_a and Z_b are interchanged, so that it can now be demonstrated that conditions I, II, and III are also sufficient to insure the physical realizability of θ_L. It already has been shown that these conditions are necessary.

Synthesis of the L Network

Consider the function $\psi(p)=(1-\theta_L)/\theta_L$. From conditions I, II, and III which both θ_L and $(1-\theta_L)$ must satisfy, it can be deduced that

$\psi(p)$ has no poles or zeros in $Re[p]>0$. (IV)
$\psi(p)$ cannot be negative real in $Re[p]>0$. That is, $0<\psi(p)$ in $Re[p]>0$. (V)

In Appendix I it is shown that conditions IV and V are the necessary and sufficient conditions for a function $\psi(p)$ to be the product of two positive real functions, and a method for obtaining the two positive real factors is given. Therefore

$$\psi(p)=(1-\theta_L)/\theta_L=X(p)Z(p)$$

where $X(p)$ is a positive real reactance function and $Z(p)$ is a positive real function.

If θ_L is the transfer function of the L network in Figure 2, then

$$(1-\theta_L)/\theta_L=Z_a/Z_b$$

Thus a network with the prescribed transfer function may be constructed by identifying $X(p)$ with $Z_a(p)$ and $1/Z_b(p)$ with $Z(p)$. Alternatively, the network may be constructed by taking $Z(p)$ as $Z_a(p)$ and $1/X(p)$ as $Z_b(p)$.

Thus it has been shown that "the necessary and sufficient conditions that a rational function, $\theta_L(p)$, of the complex frequency variable p, shall represent the transfer function of an L network are:

"$\theta_L(p)$ is real when p is real.
"$\theta_L(p)$ has no poles or zeros in $Re[p]>0$.
"$0<\theta_L(p)<1$ in $Re[p]>0$."

The Ladder Network

In the ladder network of Figure 3, it is assumed that Z_{aj} and Z_{bj} may be any physically realizable driving point impedances, but that there is no mutual coupling between any of the branches of the ladder. The fact that the transfer function of such a network may also sometimes be realized by an L network has no bearing on the following discussion.

A complete and detailed proof of some of the assertions which follow is not given here for several reasons. Chief among these is the fact that the method of synthesis which is adduced to prove the sufficiency conditions is not an attractive one from a practical point of view since it introduces a multiplicity of superfluous elements. Also, the method of synthesis

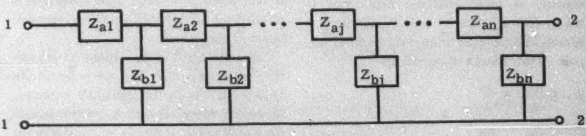


Figure 3. Ladder network

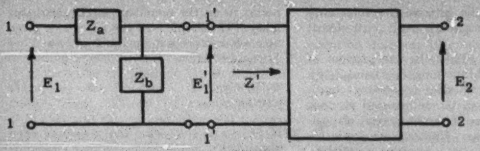


Figure 4. Cascade connection of L network and 2-terminal pair network

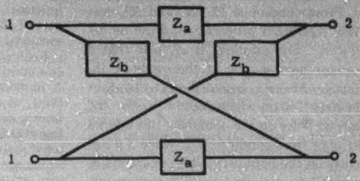


Figure 5. Symmetrical lattice network

realizes the prescribed transfer function only to within a constant factor. Obviously, there is no basic reason why both of these limitations cannot be avoided by a different synthesis procedure, and more recent investigations show promise of reaching this objective soon.

Suppose an L network with impedances Z_a and Z_b and with transfer function θ_L is connected in cascade at the left end of a 2-terminal pair network with driving point impedance Z' seen from terminals 1′ and with transfer function θ' as in Figure 4. The over-all transfer function, $\theta = E^2/E_1$, of the combined network is to be computed. By applying Thevenin's Theorem it is seen that the voltage E_1' is given by

$$E_1' = E_1 \frac{Z'}{Z' + \dfrac{Z_a Z_b}{Z_a + Z_b}} \theta_L \tag{5}$$

But since $\theta' = E_2/E_1'$, and $\theta = E_2/E_1$, we have

$$\theta = \left[\frac{Z'}{Z' + \dfrac{Z_a Z_b}{Z_a + Z_b}}\right] \theta_L \theta' \tag{6}$$

Notice that the first factor on the right of equation 6 is also a function of the θ_L class. Thus, if a prescribed transfer function is known to be the transfer function of a ladder network, it must contain as factors the transfer functions of the constituent L sections as well as other factors of the θ_L class. Generally, none of the factors of either kind appear explicitly in a given θ function when it has been reduced to its lowest terms.

With this much established, and from a knowledge of the properties of functions of the θ_L class it is evident that

θ has no poles or zeros in $Re[p]>0$. (IV)
$0<\theta<1$ when p is real and positive (VII)

and from the expression $\theta = \Delta_{12}/\Delta_{22}$, it follows that

poles of θ on the boundary are simple. (VIII)

While these are necessary conditions which θ must satisfy, it has been shown only that they are sufficient to realize θ to within a positive constant factor. In fact, to realize $K\theta$, condition VII can be disregarded. The synthesis procedure, which proves the sufficiency of conditions VI and VIII, starts with the factorization of the function $K\theta$ into the product of n factors of the θ_L class. There are a variety of ways of accomplishing this but the method which has been employed makes use of the previously proved fact that $|\arg \theta_L| \leqq \pi$ on $Re[p]=0$. For values of p on the boundary, the argument of θ is at least piece-wise continuous, having finite jump discontinuities at the poles and zeros of θ on the boundary, and has a greatest and least value. By the proper choice of a finite number of points p_i and p_j on the boundary the product $[\pi(p^2-p_i^2)/(p^2-p_j^2)][K\theta]$ can be made a function of the θ_L class while each of the inverse functions $k(p^2-p_j^2)/(p^2-p_i^2)$ is a reactive function of the θ_L class. Since the net phase displacement of θ is zero, between the origin and infinity, there are an equal number of p_i's and p_j's.

The network synthesis proceeds from right to left. Choosing any of the factors $k_1(p^2-p_{j1}^2)/(p_2-p_{i1}^2)$, it is realized, with reactive elements alone, as the first L section on the right end. The second L section from the right is obtained by choosing a second factor of the form $k_2(p^2-p_{j2}^2)/(p^2-p_{i2}^2)$. However, this is not the transfer function of the desired L section but is, rather, the transfer function of that L section when the driving point impedance of the first L section to the right is placed in parallel with the shunt impedance of the second L section. To proceed with the synthesis, it is necessary to modify the function $k_2(p^2-p_{j2}^2)/(p^2-p_{i2}^2)$ in a manner which will insure that all of the poles of the driving point impedance of the first section are included in the shunt impedance of the second section, with the proper division of the impedance obtained from the modified θ_L between the driving point impedance of the first section and the shunt arm of the second section. Such modification requires that all the poles of the transfer function of the first L section are included in the zeros of the second L section. The proof that this always can be accomplished without leaving the rest of the network unrealizable requires too much detailed consideration to present in this article.

The synthesis of the remainder of the network follows much the same pattern. It is clear that at every step in the process (except the final step) all of the poles of the driving point impedance seen to the right of an L section must be introduced as zeros of the transfer function of θ_L. Except in very unusual circumstances, this results in a multiplicity of superfluous elements. A happy circumstance is found in the final step, that of realizing the first L section on the left. In this synthesis procedure, this is the only L section of the network containing resistance, and it will be found that the transfer function of this section contains the required zeros.

The General 2-Terminal Pair Network

The transfer function of the general 2-terminal pair network always can be realized, to within a constant factor, by the symmetrical lattice structure shown in Figure 5. To prove this requires only that $\theta(p)$ satisfy the two conditions

$\theta(p)$ has no poles in $Re[p]>0$ (IX)
Poles of θ on the boundary are simple (X)

which have already been shown to be necessary, and a third condition

The real part of the residue of θ at a pole on the boundary is zero. (XI)

The necessity for condition XI follows from the fact that θ also may be represented as the ratio J_{12}/J_{11} of the familiar open circuit parameters J_{12} and J_{11} and Gewertz's[1] requirement

$$\begin{vmatrix} Re[J_{11}] & Re[J_{12}] \\ Re[J_{12}] & Re[J_{22}] \end{vmatrix} \geqq 0 \text{ when } Re[p] \geqq 0$$

Evidently $Re[J_{12}]=0$ when $Re[J_{11}]=0$ on $Re[p]=0$ and since the poles of θ on the boundary are the zeros of J_{11} it follows that J_{12} is pure imaginary at such a point. Since the derivative of J_{11} at a zero on the boundary is real and positive, condition XI is seen to hold.

From conditions IX, X, and XI, certain significant properties of θ can be deduced. For example, it is clear that θ cannot have a pole at the origin or infinity. Also, θ is bounded on any contour in $Re[p]>0$ on which $Im[\theta]=0$. For, if θ has no poles on the boundary this follows at once from condition IX, while if θ has poles on the boundary then conditions X and XI require that the Laurent expansion at such a pole, p_0, shall be $\theta(p)=ib/(p-p_0)+a_0+a_1(p-p_0)+\ldots$. In the neighborhood of p_0, θ is given approximately by $\theta \approx ib/(p-p_0)=(b/\rho)\sin\varphi+i(b/\rho)\cos\varphi$, and consequently $Im[\theta]\neq 0$ except on the boundary. Hence in $Re[p]>0$, real values of θ have an upper and lower bound and so there is a real number $K(0<K\leqq 1)$ such that $-1<K\theta<1$ whenever θ is real in $Re[p]>0$. In what follows it is assumed that $K=1$. If $K<1$, then a network with transfer function $K\theta$ is obtained.

From the preceding discussion it is clear that the functions $1\pm\theta$ have no zeros in $Re[p]>0$. Thus we can assert that the function $\psi(p)=(1-\theta)/(1+\theta)$ has no zeros or poles in $Re[p]>0$. Moreover, $\psi(p)$ cannot be negative real in $Re[p]>0$, for this requires $|\theta|>1$ in $Re[p]>0$. Consequently $\psi(p)$ is the product of two positive real functions (see Appendix I). Thus we have $(1-\theta)/(1+\theta)=Z_b/Z_a$ where Z_a and Z_b are driving point impedances.

By calculation, the transfer function of the symmetrical lattice network in Figure 5 is $\theta=(Z_a-Z_b)/(Z_a+Z_b)$ and $(1-\theta)/(1+\theta)=Z_b/Z_a$. This establishes the desired result.

This canonical form was suggested by R. M. Foster.

Appendix I

The synthesis procedure used in this paper requires the establishment of the necessary and sufficient conditions for a function, $\psi(p)$, to be the product of two positive real functions, as well as a method for obtaining the two factors. Of course, only rational fractions are encountered, and when the ψ function is not too complicated it is often possible to recognize the two positive real factors. Difficulties occur with complicated ψ functions when common factors have been cancelled, so that a systematic factoring procedure is required. Unfortunately, it appears that in the general case superfluous factors will necessarily be introduced. That is, each of the resulting positive real functions will be of higher degree in numerator and denominator than might otherwise be the case had a different method of factoring been discovered.

It turns out that with the factoring method about to be described, one of the factors is always a positive real reactance function. In this process the imaginary part of the ψ function is dealt with almost exclusively.

Let $\psi(p)=Z_1(p)Z_2(p)$ be the product of two positive real functions, $Z_1(p)$ and $Z_2(p)$. A number of properties (necessary conditions) of $\psi(p)$ can be enumerated at once from a knowledge of the properties of positive real functions. However, we are interested only in those necessary conditions which are also sufficient conditions. It is seen at once that the reciprocal function $1/\psi$ is of precisely the same class of functions as ψ since the reciprocal of a positive real function is a positive real function. A property of special significance is

$$\psi(p) \text{ is not negative real in } Re[p]>0. \text{ That is, } 0<\psi(p) \text{ in } Re[p]>0. \qquad \text{(XII)}$$

There are a number of ways of proving this statement. Perhaps it is simplest to invoke Brune's[3] condition that $|\arg Z(p)|\leqq|\arg p|$ in $Re[p]>0$. Clearly, then, $|\arg[Z_1(p)Z_2(p)]|\leqq 2|\arg p|$ in $Re[p]>0$. For $\psi(p)$ to be negative real requires $|\arg\psi|=\pi$, which is impossible when $|\arg p|<\pi/2$. This proves condition XII.

A second significant property is

$$\psi(p) \text{ has no zeros or poles in } Re[p]>0 \qquad \text{(XIII)}$$

since this is a property of positive real functions. It also might be observed that poles and zeros of $\psi(p)$ on the boundary are at most of order 2, but this is a consequence of condition XII. The proof follows from the well-known theorem which says that "a zero of multiplicity k is surrounded by $2k$ sectors of equal angles in which the real part (and also the imaginary part) of the function is alternatively positive and negative." The boundaries of the sectors for the real part are displaced from those for the imaginary part by π/k radians. Thus if $\psi(p)$ has a zero on the boundary at p_0 of order greater than 2, there is at least 1 ray from p_0 in $Re[p]>0$ on which $\psi(p)$ is negative real, which contradicts the hypothesis. A similar consideration for the reciprocal function $1/\psi(p)$, shows that poles on the boundary are at most of order 2.

The properties of $\psi(p)$ at its zeros and poles on the boundary are important here. Suppose $p_0=i\omega_0$ is a simple zero of ψ. Then in the neighborhood of p_0, ψ is given approximately by

$$\psi(p)\approx\psi'(p_0)(p-p_0)=k\rho\epsilon^{i(\vartheta+\varphi)}$$

where $\psi'(p_0)=k\epsilon^{i\vartheta}$, $(p-p_0)=\rho\epsilon^{i\varphi}$

The real and imaginary parts of ψ are approximately

$$Re[\psi]\approx k\rho\cos(\vartheta+\varphi), Im[\psi]\approx k\rho\sin(\vartheta+\varphi).$$

In $Re[p]>0$, we have $-\pi/2<\varphi<\pi/2$. Defining the principal value of ϑ by the inequality $-\pi<\vartheta\leqq\pi$, it is seen that $\psi(p)$ is real and negative in $Re[p]>0$ when $|\vartheta|>\pi/2$. Hence we must have $|\vartheta|\leqq\pi/2$, and since $Re[\psi'(p_0)]=k\cos\vartheta$ it is seen that "the real part of the derivative of ψ must be positive or zero at a simple zero on the boundary." From this it follows that "the real part of the residue of ψ at a simple pole on the boundary must be positive or zero."

In the sequel, $v(\sigma,\omega)$ will denote the imaginary part of $\psi(p)$. In particular, $v(\omega)$ or simply v will denote the imaginary part of ψ on the boundary. Evidently, for points $i\omega$ on the boundary in the neighborhood of the simple zero $p_0=i\omega_0$, $v(\omega)$ is given by $v=k(\omega-\omega_0)\cos\vartheta$, and so if $|\vartheta|\neq\pi/2$, we can assert that

$$\begin{aligned} &v<0 \text{ for } \omega<\omega_0 \\ &v>0 \text{ for } \omega>\omega_0 \end{aligned} \qquad (7)$$

If $|\vartheta|=\pi/2$, then $v(\omega_0)$ has a zero of even multiplicity.

If $p_0=i\omega_0$ is a simple pole of ψ, then

$$\begin{aligned} &v>0 \text{ for } \omega<\omega_0 \\ &v<0 \text{ for } \omega>\omega_0 \end{aligned} \qquad (8)$$

unless the real part of the residue is zero, in which case $v(\omega)$ has a zero of even multiplicity at $\omega=\omega_0$.

The behavior of ψ in the neighborhood of a double zero on the boundary is again readily investigated by means of a Taylor's series expansion. Suppose $p_0=i\omega_0$ is such a point.

Then

$$\psi(p)=\frac{\psi''(p_0)}{2!}(p-p_0)^2+\frac{\psi'''(p_0)}{3!}(p-p_0)^3+\ldots$$

and for points sufficiently close to p_0

$$\psi(p)\approx\frac{\psi''(p_0)}{2!}(p-p_0)^2$$

Writing $\psi''/2'(p_0)=k\epsilon^{i\vartheta}$, $p-p_0=\rho\theta^{i\varphi}$ gives

$$\begin{aligned} &\psi\approx k\rho^2\epsilon^{i(\vartheta-\pi)} \text{ for } \omega<\omega_0 \quad (\varphi=-\pi/2) \\ &\psi\approx k\rho^2\epsilon^{i(\vartheta+\pi)} \text{ for } \omega>\omega_0 \quad (\varphi=+\pi/2) \end{aligned}$$

and condition XII obviously requires $\vartheta=0$. Thus it is clear that at a double zero of ψ on the boundary

$$Re[\psi''(p_0)]>0 \qquad (9)$$

$$Im[\psi''(p_0)]=0 \qquad (10)$$

This statement implies that under these circumstances $v(\omega)$ must have a zero of order 3 at least. In fact, condition XII requires that $v(\omega)$ have a zero of odd order at a double zero of ψ on the boundary.

Since $1/\psi$ is also a function of the ψ class, the behavior of ψ at a double pole on the boundary is obtained from the preceding discussion since $1/\psi$ then has a double zero. It follows that at a double pole of ψ on the boundary, $v(\pi)$ has either a simple pole or a zero of odd order.

At this point it is worth while to remark that $v(\omega)$ is an odd function of the real frequency variable ω and that when $v(\omega)$ has a zero of odd multiplicity at a point for which $Re[\psi]\neq 0$, then $Re[\psi]$ is necessarily positive. This is a consequence of condition XII.

The following lemma is required in the factorization process which follows immediately.

Lemma I. Let $\psi(p)$ be a function which satisfies conditions XII and XIII. Then, if the imaginary part only of $\psi(p)$ vanishes to an odd order for some value p_0 on the boundary, the residue of the product $Y(p)\psi(p)$, where $Y(p)$ is a positive real reactance function having p_0 as a pole, will be real and positive.

The proof depends on the preceding remark that $\psi(p_0)$ must be real and positive. Since $\text{Res}[X(p)\psi(p)]_{p_0}=\text{Res}[X(p)]_{p_0}\psi(p_0,)$ and since $\text{Res}[X(p)]_{p_0}$ is real and positive, the result follows.

Factorization of ψ

In general, the factoring process consists of finding a positive real reactance function $X(p)$ such that the product $X(p)\psi(p)$ is also a positive real function. If this can be done, then clearly $\psi(p)=[X(p)\psi(p)][1/X(p)]$ and the two positive real factors have been obtained. The remainder of this Appendix contains the proof that the proper function $X(p)$ always can be found. We shall deal principally with the real function $v(\omega)$.

It is convenient to discuss the factorization under four separate conditions:

(A) When ψ has no poles or zeros on the boundary.
(B) When ψ has simple poles and zeros on the boundary.
(C) When ψ has double poles and zeros on the boundary.
(D) When $v(\omega)$ is identically zero.

Case A. Since $v(\omega)$ is an odd function, assume that it has $(2k+1)$ real zeros. It must have at least one at $\omega=0$. Thus

$$v(\omega)=\frac{\omega(\omega^2-\omega_1^2)(\omega^2-\omega_2^2)\ldots(\omega^2-\omega_k^2)}{a\omega^{2n}+b\omega^{2n-2}+\ldots}\times[g(\omega^2)]$$

Observe that if $\psi(p)=u(\sigma,\omega)+iv(\sigma,\omega)$ then on the boundary $p\psi(p)=i\omega u(\omega)-\omega v(\omega)$. Hence multiplication by p makes ωv the real part of $p\psi(p)$ and ωu the imaginary part on the boundary. Also, the sign of v is changed for all ω, and ωv is an even function of ω.

Similarly, on the boundary $(1/p)\psi(p)=(-i/\omega)u(\omega)+(1/\omega)v(\omega)$ again interchanges the roles of u and v but leaves unchanged the sign of v, and $(1/\omega)v$ is an even function.

In general, on the boundary, when p_0^2 is a real negative number, $(p_0^2=-\omega_0^2)$

$$p^2-p_0^2=-\omega^2+\omega_0^2>0 \text{ for } \omega<\omega_0$$
$$p^2-p_0^2=-\omega^2+\omega_0^2<0 \text{ for } \omega>\omega_0$$

and similarly for the reciprocal function $1/(p^2-p_0^2)$. Thus $(-\omega^2+\omega_0^2)h(\omega)$ has the same sign as $h(\omega)$ for $\omega<\omega_0$ and has the opposite sign for $\omega>\omega_0$. Identical results obtain for $[1/(-\omega^2+\omega_0^2)][h(\omega)]$ deleting the point $\omega=\omega_0$ at which the product may become infinite.

Returning to the function $v(\omega)$, it will be assumed that the k factors $(\omega^2-\omega_1^2)$, $(\omega^2-\omega_2^2)$, and so forth, are distinct. This is not an essential restriction, for if $(\omega^2-\omega_j^2)$ is a factor of even multiplicity it does not affect the sign of $v(\omega)$ and is ignored. If $(\omega^2-\omega_j^2)$ is a factor of odd multiplicity, it is treated as the product of a factor of even multiplicity and a simple factor. By hypothesis, v has no poles on the boundary, hence it is bounded in absolute value and has variations in sign equal to the number of simple and odd multiple factors for $\omega>0$.

The factoring process starts by considering the sign of v between $\omega=0$ and $\omega=\omega_1$. If v is positive, multiply ψ by $1/p$. Hence $(1/\omega)v$ remains positive in the interval and is an even function. If v is negative between $\omega=0$ and $\omega=\omega_1$, multiply ψ by p. The function $-\omega v$ now has the opposite sign of v everywhere for $\omega>0$ and is an even function.

In either case, $(1/\omega)v$ or $-\omega v$ is negative between ω_1 and ω_2. Multiply $(1/p)\psi$ by $(p^2-p_1^2)$ or, if the other alternative obtains, multiply $p\psi$ by $1/(p^2-p_1^2)$. Such multiplication does not affect the sign of $(1/\omega)v$ or $-\omega v$ between $\omega=0$ and $\omega=\omega_1$ but changes the sign of the result for all $\omega>\omega_1$. Hence $(-\omega^2+\omega_1^2)(1/\omega)v$ or $-1/(-\omega^2+\omega_1^2)\omega v$ is positive between $\omega=0$ and $\omega=\omega_2$. Note also that multiplication by $(-\omega^2+\omega_1^2)$ results in a zero of even multiplicity while multiplication by $1/(-\omega+\omega_1)$ simply cancels a factor leaving the result regular at $\omega=\omega_1$.

Between ω_2 and ω_3 the functions $(-\omega^2+\omega_1^2)(1/\omega)v$ and $-[1/(-\omega^2+\omega_1^2)][\omega v]$ are negative. Multiplication by $1/(-\omega^2+\omega_2^2)$ and $(-\omega^2+\omega_2^2)$ respectively makes the resulting functions positive in the interval $0<\omega<\omega_3$. Since the number of variations of sign of the function $v(\omega)$ is finite, then by proceeding in this fashion, alternating the zeros and poles, the process eventually will terminate, resulting in the function, taking the first alternative

$$\frac{1}{\omega}\frac{(-\omega^2+\omega_1^2)(-\omega^2+\omega_3^2)\ldots}{(-\omega^2+\omega_2^2)(-\omega^2+\omega_4^2)\ldots}v$$

which is positive for all ω.

The original function $\psi(p)$ now has become

$$\frac{1}{p}\frac{(p^2-p_1^2)(p^2-p_3^2)\ldots}{(p^2-p_2^2)(p^2-p_4^2)\ldots}\psi(p)=Z(p)$$

Hence

$$\psi(p)=Z(p)\left[p\frac{(p^2-p_2^2)(p^2-p_4^2)\ldots}{(p^2-p_1^2)(p^2-p_3^2)\ldots}\right]$$

By the reactance theorem, the second factor on the right

$$p\frac{(p^2-p_2^2)(p^2-p_4^2)\ldots}{(p^2-p_1^2)(p^2-p_2^2)}=X(p)$$

is a positive real reactance function.

Likewise, the function

$$Z(p)=\frac{1}{p}\frac{(p^2-p_1^2)(p^2-p_3^2)\ldots}{(p^2-p_2^2)(p^2-p_4^2)\ldots}\psi(p)$$

is a positive real function for: (1) $Z(p)$ has no poles in $Re[p]>0$, (2) $Re[Z(i\omega)]\geq 0$, (3) poles on the boundary are simple and have positive real residues. This follows from the fact that $Z(p)=[1/X(p)]\psi(p)=Y(p)\psi(p)$. But $Y(p)$ is a positive real reactance function whose poles occur only when $v(\omega)$ has zeros of odd multiplicity. Hence by Lemma I, the residues at those poles are positive real constants.

Case B. $\psi(p)$ has simple poles and zeros on the boundary.

It has been shown in equation 8 that at a simple pole p_0 of ψ on the boundary, $v(\omega)$ is positive in the immediate neighborhood $\omega<\omega_0$ and negative in the immediate neighborhood $\omega>\omega_0$, or else has a zero of even multiplicity at ω_0, in which case that pole and the following discussion is ignored. Clearly there must be an odd number of zeros of odd multiplicity (hereafter called critical zeros) of $v(\omega)$ between each pole on the boundary. Because $v(\omega)$ is an odd function, it must have either a critical zero or a pole at the origin.

If $v(\omega)$ has a pole at the origin, then it must be negative for values of $\omega>0$ in the immediate neighborhood. Hence the first multiplicative factor is $(-\omega)$, the next $1/(-\omega^2+\omega_1^2)$, the next $(-\omega^2+\omega_2^2)$, and so forth, until the first pole of $v(\omega)$ located at ω_j is reached. Notice that factors with even subscripts appear in the numerator, those with odd subscripts in the denominator. From the discussion in the preceding paragraph it is clear that the subscript j must be even. Hence the factor $(-\omega^2+\omega_j^2)$ appears in the numerator and is precisely the factor required to remove the pole at ω_j. As applied to the ψ function, the factor $(p^2-p_j^2)$ removes the simple pole at p_j. It is clear also that the same remarks apply to any interval of the boundary between two poles. The completion of this process produces the function

$$p\frac{(p^2-p_2^2)(p^2-p_4^2)\ldots(p^2-p_k^2)}{(p^2-p_1^2)(p^2-p_3^2)\ldots(p^2-p_{k-1}^2)}\psi(p)=[1/X(p)]\psi(p)=Z(p)$$

or

$$\psi(p)=X(p)Z(p)$$

Obviously, $X(p)$ is a positive real reactance function. As in case A, (1) $Z(p)$ has no poles in $Re[p]>0$, and (2) $Re[Z(i\omega)]\geq 0$. If $\psi(p)$ has no poles on the boundary of the "ignorable" variety previously mentioned, then by Lemma I $Z(p)$ has positive real residues at all its poles on the boundary and is a positive real function. The residue of $Z(p)$ at an ignorable pole of $\psi(p)$ is also positive real. For at such a point, say p_k, the real part of the residue of $\psi(p)$ is zero and for points in $Re[p]>0$ sufficiently close to p_k, $Im[\psi]\approx(b/\rho)\cos\varphi$ $(-\pi/2<\varphi<\pi/2)$ where b is the imaginary part of the residue. Clearly, the sign of $v(\omega_k)$ near ω_k is the same as the sign of b. Since ω_k is not a critical point of v, then in any event $1/X(p_k)$ is regular. Suppose, for instance, $v(\omega_k)<0$ near ω_k, implying $b<0$. If there are an even number of critical points between ω_k and $\omega=0$, then $v(\omega)$ is negative for $\omega>0$ in the vicinity of the origin, and the factorization starts with p, whence $1/X(p_k)=ia$ $(a>0)$. If the number of critical points is odd, then the factorization starts with $1/p$ and again $1/X(p_k)=ia$. Thus the residue at p_k of $[1/X(p)]\psi(p)$ is $(-ib)(ia)$ which is real and positive. A similar analysis for $v(\omega)>0$ near ω_k again yields a positive real residue. Thus, in any case, $Z(p)$ is a positive real function.

If $v(\omega)$ has a zero at the origin it may be either positive or negative for values of $\omega>0$ in the immediate neighborhood of the origin. If $v(\omega)$ is positive there must be an even number of critical zeros between $\omega=0$ and $\omega=\omega_j$ (j odd), the first pole of $v(\omega)$. The factoring process is started by multiplying ψ by $1/p$, factors with odd subscripts appearing in the numerator so that again the factor $(p^2-p_j^2)$ appears in the numerator to cancel the pole of ψ at p_j. From this point the process proceeds as before.

If $v(\omega)$ is negative for values of $\omega>0$ in the immediate neighborhood of the origin, the discussion given for the case when v has a pole at the origin applies word for word. Thus, in either case one of the following possibilities must be obtained.

$$p\frac{(p^2-p_2^2)(p^2-p_4^2)\ldots(p^2-p_k^2)}{(p^2-p_1^2)(p^2-p_3^2)\ldots(p^2-p_{k-1}^2)}\psi(p)=B(p)\psi(p)=Z(p)$$

or

$$\frac{1}{p}\,\frac{(p^2-p_1^2)(p^2-p_3^2)\ldots(p^2-p_{k-1}^2)}{(p^2-p_2^2)(p^2-p_4^2)\ldots(p^2-p_k^2)}\psi(p) = B(p)\psi(p)=Z(p)$$

whence $\psi(p)=[1/B(p)]Z(p)=X(p)Z(p)$ where, as before, $X(p)$ is a positive real reactance function and $Z(p)$ is a positive real function.

Case C. $\psi(p)$ has double poles and zeros on the boundary.

It has been shown, following equation 10, that at a double zero of ψ on the boundary, $v(\omega)$ must have a zero of odd order, the multiplicity being not less than 3. Obviously such a zero is a critical zero. From the Taylor's series expansion preceding equation 9 it is clear that in the vicinity of the double pole the imaginary part of ψ is given approximately by $Im[\psi]\approx k\rho^2 \sin 2\varphi$, so that as φ varies from $-\pi/2$ to $+\pi/2$, $Im[\psi]$ goes from negative to positive values. In other words, in the neighborhood of the double zero, p_0, $v(\omega)$ is negative for $\omega<\omega_0$ and positive for $\omega>\omega_0$. Hence the factor to be inserted corresponding to such a point always must be introduced into the denominator.

Similarly, at a double pole, p_0, on the boundary $v(\omega)$ has either a simple pole or a zero of odd order. Thus, such a point is also a critical point of v, and it follows from the preceding that in the neighborhood of this point $v(\omega)$ is positive for $\omega<\omega_0$ and negative for $\omega>\omega_0$. Hence the factor to be inserted corresponding to this point always must be introduced into the numerator. The arguments which show that the factors which result from the alternation process will actually occur in denominator and numerator as required are the same as for case B. Thus the factorization proceeds exactly as for case B, and we will then obtain $\psi(p)=X(p)Z(p)$ where $X(p)$ and $Z(p)$ will contain one each of the double poles and zeros.

It should be observed that in case C, $Z(p)$ may contain poles on the boundary arising from circumstances not covered by cases A and B. These are the poles corresponding to the double poles of $\psi(p)$. Such a pole of $Z(p)$ occurring, say, at p_k is, of course, a simple pole. The residue of $Z(p)$ at p_k can be inferred from the properties of the derivative of $1/Z(p)$ at that point. Writing $1/Z(p)=X(p)[1/\psi(p)]$ it is seen that $X(p)$ has a simple pole at p_k while $1/\psi(p)$ has a double zero there. Expanding each factor of the right member in a Laurent and Taylor series about p_k yields the expression

$$\frac{1}{Z(p)}=\left[\frac{b_1}{p-p_k}+a_0+\ldots\right]\times [a_2(p-p_k)^2+a_3(p-p_k)^3+\ldots]$$

The Laurent and Taylor series, being both uniformly convergent in a sufficiently small neighborhood of p_k, assures that the product series will be the series expansion of $1/Z(p)$. Thus $1/Z(p)=b_1a_2(p-p_k)+(b_1a_3-a_0a_2)(p-p_k)^2+\ldots$ Since b_1 is the residue of $X(p)$ at p_k, $b_1>0$. Since $a_2=Re[\psi''(p_k)]$, then by equation 9 $a_2>0$, and it follows that the derivative of $1/Z(p)$ at p_k is positive real. Hence the residue of $Z(p)$ at the simple pole p_k is positive real. Thus again under all circumstances $Z(p)$ is a positive real function.

Case D. The function $v(\omega)$ is identically zero.

In effect, this is the simplest factorization of all. If $\psi(p)$ is real on the boundary, then by the Schwarz Reflection Principle[4] it must have all of its poles and zeros on the boundary. For if $\psi(p)$ has a zero in $Re[p]<0$ it must have a corresponding zero in $Re[p]>0$. By hypothesis $\psi(p)$ has no zeros in $Re[p]>0$ and hence has no zeros in $Re[p]<0$. Treating the function $1/\psi(p)$ similarly yields the conclusion that $\psi(p)$ has no poles in $Re[p]<0$. Thus if $\psi(p)$ is not a constant, all of its poles and zeros are on the boundary. As previously shown, poles and zeros on the boundary are at most of order 2. More than that, however, by applying the Argument Principle and condition XII, it is seen that no more than two poles or zeros can occur in succession, each counted in accordance with its multiplicity.

Since $\psi(i\omega)$ is real it is also an even function of ω and so zeros and poles at the origin or at infinity must be of order 2. With this it is clear that $\psi(p)$ always may be written in one of three forms, (1) $\psi(p)=K(p/p)g(p)$, (2) $\psi(p)=Kp^2g(p)$, (3) $\psi(p)=K(1/p^2)g(p)$, where K is a positive constant and $g(p)$ with its reciprocal is regular at the origin and infinity. Evidently the numerator and denominator of $g(p)$ are of the same degree and so $g(p)$ must have the same number of poles as zeros, each counted in accordance with its multiplicity. Consequently, the poles and zeros of $g(p)$ must occur in pairs although not necessarily in order. But since not more than two poles or zeros may occur in succession, it is always possible to factor $g(p)$ into the product of two functions, each having simple poles and zeros on the boundary, the poles and zeros alternating in each factor separately.

The remainder of the factoring process consists merely in assigning the factors (1) Kp and $1/p$, (2) Kp and p, (3) K/p and $1/p$ properly to the two previously described factors of $g(p)$. That this can always be done is readily seen when it is observed that in (1) the first critical point (zero or pole) must be simple, in (2) the first critical point must be a pole, and in (3) the first critical point must be a zero. Thus the function $\psi(p)$ always can be factored into the product of two positive real reactance functions, $\psi(p)=X_1(p)X_2(p)$.

References

1. SYNTHESIS OF A FINITE FOUR-TERMINAL NETWORK, A. Gewertz. *Journal of Mathematics and Physics*, Massachusetts Institute of Technology (Cambridge, Mass.), volume 12, 1932–33, pages 1–257.

2. SYNTHESIS OF A FINITE TWO-TERMINAL NETWORK WHOSE DRIVING POINT IMPEDANCE IS A PRESCRIBED FUNCTION OF FREQUENCY, O. Brune. *Journal of Mathematics and Physics*, Massachusetts Institute of Technology (Cambridge, Mass.), volume 10, October 1931, pages 191–235.

3. A REACTANCE THEOREM, R. M. Foster. *Bell System Technical Journal* (New York, N. Y.), April 1924, pages 259–67.

4. THE THEORY OF FUNCTIONS (book), E. C. Titchmarsh. Oxford University Press, London, England, 1939.

No Discussion

Design of Low-Frequency Constant Time Delay Lines

CLIFFORD M. WALLIS
FELLOW AIEE

DELAY lines are used in many applications in both high-frequency and low-frequency fields. In low-frequency applications such lines consist of a cascade connection of inductance-capacitance network sections employing lumped constants. Ideally, the line should have no attenuation; its characteristic impedance should be constant, and the phase shift should be a linear function of frequency. With the latter condition, the time delay is constant and the character of the signal is unchanged in its transmission over the line. In design practice, one attempts to meet as closely as possible these ideal conditions within the limits of cost, space, and practicability.

When the desired delay time is large, that is, in the magnitude of milliseconds and the passband ranges from zero frequency to several kilocycles, the line necessarily must be long in terms of the number of sections used. In addition to the cost, space requirements, line loss, and simplicity of construction become important factors of consideration. Generally speaking, the closer the tolerance on constancy of delay time the smaller must be the delay time per section and hence the greater the number of sections. This means larger line loss, more space, and higher cost for a fixed over-all delay time. Again, to obtain constancy of delay time over a large portion of the passband by use of a complicated network section increases the space required and line cost. Consequently, one must determine the kind of network that best meets the electrical, space, and economic requirements of the application. It is the purpose of this paper to present data and to compare the relative merits of several types of networks that are particularly useful where essentially constant delay for signals ranging in frequency from near zero to some arbitrary upper limit is required.

Where the frequency range lies in this part of the spectrum an inductance-capacitance network of the low-pass type is employed. There are a number of these whose characteristics make them useful in delay line design. In particular, they are:

The M-derived section.
The bridged-T section.
The capacitance shunted section.
The unsymmetrical Pierce section.

Analyses of the first three have been presented in a recent article by Hebb, Horton and Jones.[1] Although these circuits often are used in filter design[2] their time-delay characteristics are not as commonly known. The unsymmetrical Pierce section is perhaps even less familiar. As far as the author knows, its virtue as a delay network was first discovered by G. W. Pierce around 1919, though little mention of it has been made in the literature. The major portion of this discussion will be devoted to presenting an analysis and the characteristics of this circuit. For the sake of comparison, the time-delay characteristics and design equations of the M-derived and bridged-T sections will be included. The latter two symmetrical sections and the Pierce unsymmetrical section are shown in Figures 1(A), 1(B), and 1(C) respectively. All are of the T-type structure employing mutual coupling between the inductive elements of the T.

The following symbols are used in the analysis:

Z_0 = characteristic impedance
Z_{00} = characteristic impedance at $f \to 0$
$\omega = 2\pi f$
β = phase-shift angle
f_0 = frequency at which $\beta = 180$ degrees
T_g = group delay time $= d\beta/d\omega$
T_0 = delay time at $f \to 0$ (called zero-frequency delay time)
k = coupling coefficient
Z_1Z_2 = iterative impedances of unsymmetrical section in forward and backward direction

Group delay time represents the delay time for a narrow band of frequencies centered about the angular frequency (ω). Defined in this manner it is the slope of the phase-shift curve.

The conventional assumption is made that the coils and capacitors are lossless components as far as determining the delay and impedance characteristics of the section.

In designing a delay line one usually starts with the following factors specified to meet a given application:

1. Frequency range.
2. Total delay time or delay time between feed-in points if signals are to be injected into the line at several points.
3. Nominal (zero frequency) characteristic impedance.
4. Allowable variation (tolerance) of items 2 and 3 over the specified frequency range.

It is convenient to express the delay time in terms of its zero frequency value and the per-cent variation. Therefore, the design equations are expressed in terms of this parameter, the nominal characteristic impedance, and such other parameters as are needed.

M-Derived Section

A type of M-derived section which offers an optimum time-delay characteristic in its passband is one where the two coils are connected series aiding. Its equivalent circuit is shown in Figure 2. In terms of the zero-frequency characteristic impedance and delay time, the constants of the section are

---

Paper 52-150, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE South West District Meeting, St. Louis, Mo., April 15–17, 1952. Manuscript submitted January 11, 1952; made available for printing March 3, 1952.

CLIFFORD M. WALLIS is at the University of Missouri, Columbia, Mo.

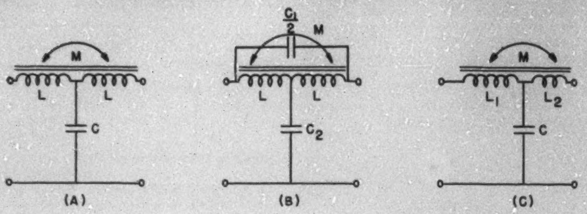


Figure 1. Three inductance-capacitance networks used in low-frequency time-delay lines

$$L = \frac{T_0 Z_{c0}}{2(1+k)} \qquad (1)$$

$$C = \frac{2L(1+k)}{Z_{c0}^2} = \frac{T_0}{Z_{c0}} \qquad (2)$$

The cutoff frequency and phase-shift angle are

$$f_c = \frac{m}{\pi T_0} \qquad (3)$$

$$\beta = 2 \sin^{-1}\left[\frac{mu}{\sqrt{1-(1-m^2)u^2}}\right] \qquad (4)$$

where $m = \sqrt{(1+k)/(1-k)}$ and $u = f/fc$. The group delay time at any frequency within the passband is

$$T_g = \frac{m}{\pi f_c \sqrt{1-u^2}\,[1-(1-m^2)u^2]} \qquad (5)$$

Normalizing with respect to T_0.

$$\frac{T_g}{T_0} = \frac{1}{\sqrt{1-u^2}[1-(1-m^2)u^2]} \qquad (6)$$

Curves of $(T_g - T_0)/T_0$ are plotted in Figure 3 for several values of coupling coefficient. Included is the one for $k=0$, representing the constant-K type of section. For values of k in the neighborhood of 0.25 we see that the delay time remains essentially constant over approximately 60 per cent of the passband. The serious disadvantage of this type of section lies in the difficulty of obtaining a coupling coefficient this low when the two coils are wound on a common toroidal core of magnetic material. To obtain a value of $k=0.25$, and at the same time maintain high coil Q's, would demand a special arrangement of coils.

Bridged-T Section

Though requiring an extra component, the bridged-T circuit has better delay characteristics for high values of k than the M-derived section. The equivalent electric circuit is shown in Figure 4 for the coils connected series aiding. The phase shift in the passband is

$$\beta = 2 \tan^{-1}\left[u_2 \sqrt{\frac{1+k}{1-k}} \times \frac{1}{\sqrt{1-(u_1^2+u_2^2)+u_1^2 u_2^2}}\right] \qquad (7)$$

where

$u_1 = f/f_1$
$u_2 = f/f_2$

and

$$f_1 = \frac{1}{2\pi\sqrt{C_1 L(1+k)}} \qquad (8A)$$

$$f_2 = \frac{\sqrt{2}}{2\pi\sqrt{C_2 L(1-k)}} \qquad (8B)$$

For the characteristic impedance we write

$$Z_c = \sqrt{\frac{2L(1+k)}{C_2}}\sqrt{\frac{1-(u_1^2+u_2^2)+u_1^2u_2^2}{(1-u_1^2)^2}} \qquad (9)$$

From equation 7 one may obtain the expression for the delay time ratio as

$$\frac{T_g}{T_0} = \frac{1-u_1^2u_2^2}{\left[1-u_1^2+\left(\frac{2k}{1-k}\right)u_2^2+u_1^2u_2^2\right] \times \sqrt{1-(u_1^2+u_2^2)+u_1^2u_2^2}} \qquad (10)$$

In general there are two frequencies at which the phase-shift angle β is equal to 180 degrees, indicating the presence of two passbands. These frequencies are f_1 and f_2. If we let $a = f_1^2/f_2^2 = u_2^2/u_1^2$, equations 9 and 10 become

$$Z_c = \sqrt{\frac{2L(1+k)}{C_2}}\sqrt{\frac{1-au_1^2}{1-u_1^2}} \qquad (11)$$

and

$$\frac{T_g}{T_0} = \frac{1-au_1^4}{\left[1+u_1^2\left(\frac{2ka}{1-k}-1\right)+au_1^4\right] \times \sqrt{1-au_1^2)(1-u_1^2)}} \qquad (12)$$

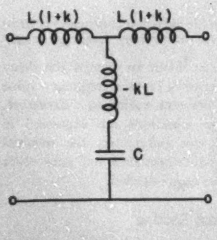


Figure 2 (above). Equivalent circuit of the M-derived section shown in Figure 1(A)

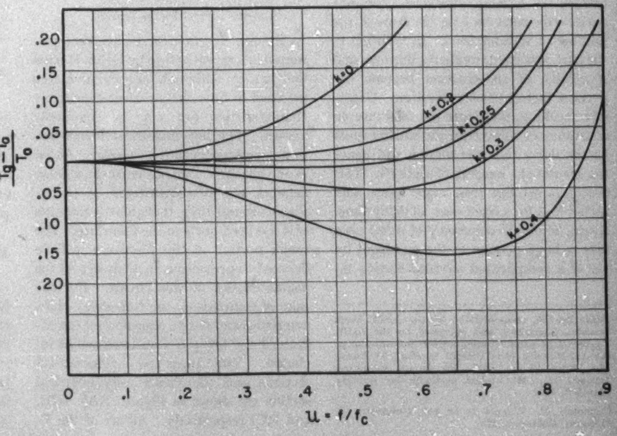


Figure 3 (right). Per cent variation of delay time versus frequency ratio for M-derived section of Figure 1(A)

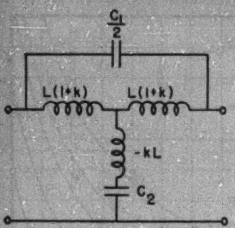


Figure 4 (above). Equivalent circuit of bridged-T section shown in Figure 1(B)

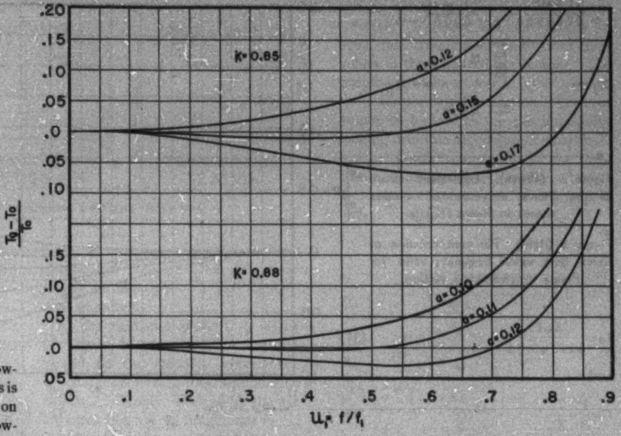


Figure 5 (right). Per cent variation of delay time versus frequency ratio for bridged-T section of Figure 1(B)

A type of inductor widely used in low-frequency (1 to 50 kilocycles) delay lines is one which consists of two coils wound on a toroid core. The core is a bonded low-loss magnetic material such as molybdenum permalloy. The coil Q is relatively high and, as there is little stray field, the coil units can be arranged in a compact manner. Their coupling coefficient k ranges from about 0.8 to 0.9 and for the smaller sized units centers pretty well around 0.87. With a value of k in this range the parameter a is less than unity and f_1 is the lower cutoff frequency. The curves of Figure 5 show that for a minimum variation of delay time the values of a range between 0.11 to 0.15, depending on the exact value of k.

When the allowable variation in the delay time of a line for a given frequency range is specified, along with the line impedance, reference to a family of curves such as those of Figure 5 allows one to determine the proper value of a. With this parameter determined, f_1 and f_2 are fixed and the constants of the section as well as the delay time per section can be calculated from the equations

$$T_o=\frac{1}{\pi f_2}\sqrt{\frac{1+k}{1-k}} \qquad (13)$$

$$C_2=\frac{T_o}{Z_{co}} \qquad (14)$$

$$L=\frac{T_oZ_{co}}{2(1+k)} \qquad (15)$$

The capacitance C_1 is determined from equation 8(A).

Pierce Unsymmetrical Section

The Pierce unsymmetrical section is shown in Figure 1(C). The two coils, whose self-inductances are designated as L_1 and L_2, are connected series aiding, and are wound to have a coefficient of coupling as close to unity as possible. If the ratio of L_1 to L_2 is made large and the sections are connected on an iterative basis, it is possible to secure a delay time that is nearly constant over a large portion of the passband. The avoidance of the need for loose coupling between the two coils is a distinct advantage of the Pierce section over the symmetrical M-derived type.

When connected in an iterative manner all the sections are oriented in the same direction. If each end of the line is terminated in an impedance that takes the place of an infinite continuation of the line, it is said to be terminated on an iterative basis. These impedances, designated as Z_1 and Z_2, are called the iterative impedances of the section.[8]

The equivalent circuit is shown in Figure 6. In terms of the circuit constants the iterative impedances are

$$\left.\begin{matrix}Z_1\\Z_2\end{matrix}\right|=\frac{1}{2}\times\left\{\sqrt{(L_1+L_2+2M)[4S-\omega^2(L_1+L_2-2M)]}\pm j(L_1-L_2)\omega\right\} \qquad (16)$$

It may be noted that within the passband Z_1 and Z_2 are complex conjugate impedances. For frequencies in this region the phase-shift angle can be expressed either as

$$\beta=2\sin^{-1}\left[\frac{\omega}{2}\sqrt{\frac{L_1+L_2+2M}{S+M\omega^2}}\right] \qquad (17A)$$

or

$$\beta=2\tan^{-1}\left[\omega\sqrt{\frac{L_1+L_2+2M}{4S-\omega^2(L_1+L_2-2M)}}\right] \qquad (17B)$$

For the delay time we may write

$$T_g=\left[\frac{S}{S+M\omega^2}\right]\times\sqrt{\frac{4(L_1+L_2+2M)}{4(S+M\omega^2)-\omega^2(L_1+L_2+2M)}} \qquad (18)$$

There is one other factor of interest, particularly for applications where signals are to be injected into the line at several points. This is the impedance looking into the section between ground side and the point of the capacitor connection between coils. When the section is terminated on the iterative basis, this impedance is largely resistive for frequencies within a major portion of the passband. Hence, this is the point where signals are most conveniently fed into the line if bridging generators are used. This impedance, denoted by Z_T in Figure 7, is

$$Z_T=\left[\frac{S^2}{S+M\omega^2}\right]\sqrt{\frac{L_1+L_2+2M}{4S-\omega^2(L_1+L_2-2M)}}-j\omega\frac{MS}{S+M\omega^2} \qquad (19)$$

The zero-frequency values of the several impedances and of the delay time are evident from examination of the preceding equations. In particular it may be seen that $2Z_{T0}=Z_{10}=Z_{20}=Z_{co}$.

Again it is convenient to set up the design equations in terms of the zero-frequency values of the delay time and iterative impedance, the coupling coefficient $k=M/\sqrt{L_1L_2}$, and the parameter r, which is the ratio L_1/L_2. Expressed in this way

$$f_c=\frac{1}{\pi T_o}\sqrt{\frac{r+2k\sqrt{r}+1}{r-2k\sqrt{r}+1}} \qquad (20)$$

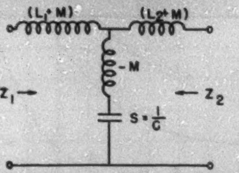
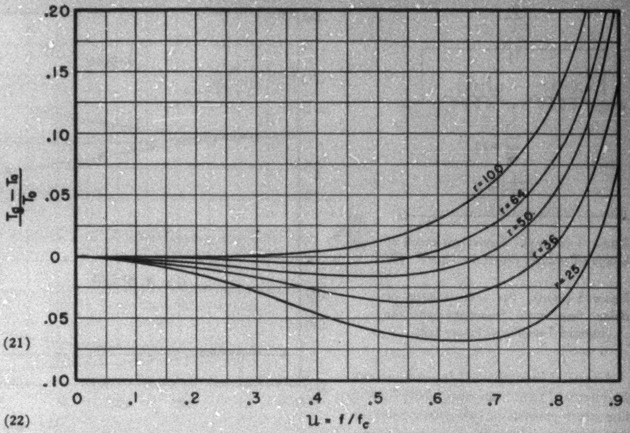


Figure 6 (above). Equivalent circuit of the Pierce unsymmetrical section shown in Figure 1(C)

Figure 8 (right). Per cent variation of delay time versus frequency ratio for Pierce unsymmetrical section



$$L_2=\frac{Z_{co}T_o}{r+2k\sqrt{r}+1} \tag{21}$$

and

$$C=\frac{1}{S}=\frac{T_o}{Z_{co}} \tag{22}$$

The delay time, normalized with respect to T_o, is

$$\frac{T_g}{T_o}=\frac{r-2k\sqrt{r}+1}{[r-2k\sqrt{r}+1+4k\sqrt{r}u^2]\sqrt{1-u^2}} \tag{23}$$

where $u=f/fc$. The impedance expression becomes

$$\left.\begin{matrix}Z_1\\ Z_2\end{matrix}\right|=Z_{co}\left[\sqrt{1-u^2}\pm j\frac{(r-1)u}{\sqrt{r^2+r(2-4k^2)+1}}\right] \tag{24}$$

Minimum variation of delay is secured when r is large and k approaches unity. For values of $k\geq 0.95$ the optimum values of r are above 50. With this magnitude very little error is made in simplifying the preceding equations by assuming $k=1$. Whence

$$f_c=\frac{1}{\pi T_o}\left[\frac{\sqrt{r}+1}{\sqrt{r}-1}\right] \tag{25}$$

$$L_2=\frac{Z_{co}T_o}{(\sqrt{r}+1)^2} \tag{26}$$

$$\frac{T_g}{T_o}=\frac{(\sqrt{r}-1)^2}{[4\sqrt{r}u^2+(\sqrt{r}-1)^2]\sqrt{1-u^2}} \tag{27}$$

for $k\geq 0.9$, $r\geq 50$ (25)–(27)

Expressions for the iterative and bridging impedances, normalized with respect to Z_{co} become

$$\left.\begin{matrix}Z_1/Z_{co}\\ Z_2/Z_{co}\end{matrix}\right|=\sqrt{1-u^2}\pm ju \tag{28}$$

$$\frac{2Z_T}{Z_{co}}=\frac{1}{\left[1+\frac{4\sqrt{r}}{(\sqrt{r}-1)^2}u^2\right]\times\sqrt{1-u^2}}-j\frac{u}{\frac{r-1}{4\sqrt{r}}+\left(\frac{\sqrt{r}+1}{\sqrt{r}-1}\right)u^2} \tag{29}$$

For $k\geq 0.95$, $r\geq 50.0$ (28)–(29)

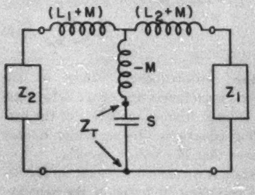


Figure 7 (above). Equivalent circuit to determine the shunt input impedance Z_r

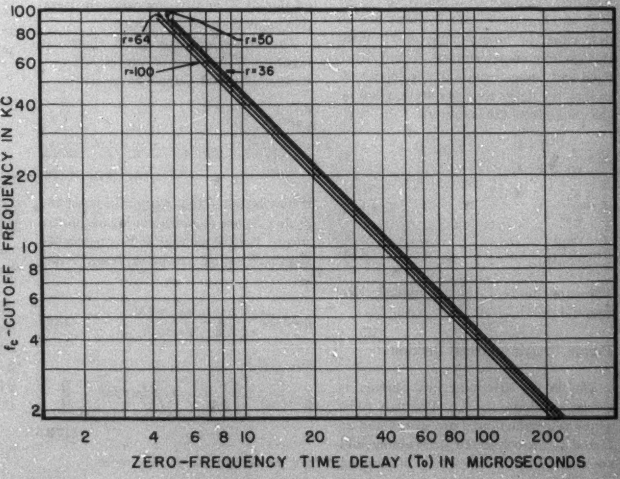


Figure 9 (right). Cutoff frequency versus zero-frequency delay time

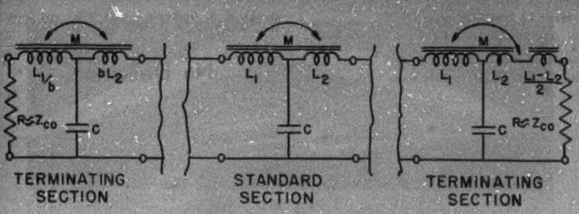


Figure 10. A method of terminating the Pierce delay line

Values of $(T_g-T_o)/T_o$ are plotted as functions of u in Figure 8 for several values of the parameter r. It may be seen that with a value of $r=64$, for example, the delay time is constant to within ±1 per cent over 60 per cent of the passband.

Equation 28 shows that the scalar magnitude of either iterative impedance is equal to Z_{co} throughout the passband; only the resistive and reactive components vary. Figure 9 shows plots of cutoff frequency f_c versus the delay time T_o for several values of r. Using log-log scales gives straight line plots that are useful for quick reference.

Termination

Terminating delay lines that use symmetrical sections presents no serious difficulty. A resistance is chosen which best matches the characteristic impedance over the operating frequency range. However, the fact that the Pierce unsymmetrical section has iterative impedances that are complex conjugates somewhat complicates the problem of securing proper termination of a line using this type of section.

At the right end of the line, referring to Figure 10, the terminal impedance should be the iterative impedance Z_1 which is inductive. Proper termination in this case is simply obtained by the use of an extra coil whose inductance is $(L_1-L_2)/2$ and a resistance that gives the best over-all match to the resistive component of Z_1. In Figure 10 the extra coil is shown as part of the terminating section on the right.

On the left end of the line, the terminal impedance should be Z_2 which has a negative reactive component $-j\omega(L_1-L_2)/2$. Such a reactance cannot be produced by a simple circuit. Over a limited frequency range it can be simulated fairly well by use of a parallel resistance-capacitance network with proper selection of components. Better termination is secured by the method indicated in Figure 10. The last section is terminated in Z_{co} and its coil inductances L_1 and L_2 are respectively decreased and increased by an equal ratio (designated as b) from their values in the standard line section. The coupling is made as near unity as possible as in the standard line section. Then for values of b close to 2.0 the impedance looking into this section approximates Z_2 over a considerable portion of the passband. The equations for the resistive and reactive components of this impedance normalized with respect to Z_{co} are

$$\frac{R}{Z_{co}}=\frac{\left[\frac{(\sqrt{r}-1)^2}{4\sqrt{r}}+u^2\right]^2}{\frac{(r-1)^2}{4r}u^2+\frac{r}{4}\left[\frac{2}{b}u^2-\frac{(\sqrt{r}-1)^2}{2r}\right]^2} \qquad (30)$$

$$\frac{X}{Z_{co}}=-ju\left[\frac{\sqrt{r}-1}{\sqrt{r}+1}\right]\times$$

$$\frac{\left[\frac{(\sqrt{r}-1)^2}{8r}\left(r+1-b-\frac{r}{b}\right)\right]+\left[\frac{r}{2b}+\frac{1}{2}+\frac{\sqrt{r}}{b}-\frac{b}{2r}(\sqrt{r}+1)^2\right]u^2}{\frac{(r-1)^2}{4r}u^2+\frac{r}{4}\left[\frac{2}{b}u^2-\frac{(\sqrt{r}-1)^2}{2r}\right]^2} \qquad (31)$$

where r is the ratio $-L_1/L_2$ used in the standard line section.

The curves of Figure 11 show the degree of approximation to Z_2 that can be realized with such a terminating section when $r=64$. A value of $b=2.0$ appears to give optimum matching considering both resistive and reactive components. Lines built with this termination have shown remarkably good characteristics, that is, constancy of delay time and smooth line operation.

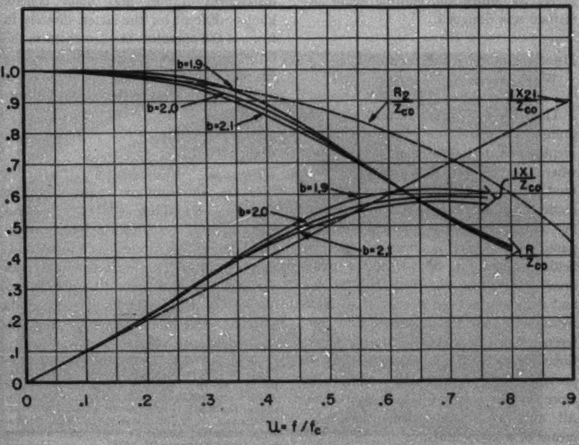


Figure 11. Curves showing degree of matching obtained by use of terminating section on left end of Pierce line as shown in Figure 10

Conclusion

In an application where the required delay time is large, necessitating a line of many sections, and space is at a premium, the Pierce section offers distinct advantages. Only two components per section are needed—one capacitor and a double-winding inductor. The section constants are of sizes that are procured easily in practice. For example, for a 1,000-ohm 50-microsecond section, $C=0.05$ microfarad, $L_2=0.61$ millihenry. With $r=64$, $L_1=39.5$ millihenrys.

Its time-delay characteristic compares very favorably with the other types mentioned. Although the bridged-T and capacitor-shunted types can be made to give constant delay over a slightly greater percentage of the passband, each requires an extra capacitor. For constant time delay the M-derived type demands a low coupling coefficient between coils of the inductor. This is not easily obtained where magnetic cores are used.

References

1. On Design of Networks for Constant Time Delay, M. H. Hebb, C. W. Horton, F. B. Jones. *Journal of Applied Physics* (New York, N. Y.), volume 20, June 1949, pages 616–20.

2. Electric Circuits and Wave Filters (book), A. T. Starr. Sir Isaac Pitman and Sons, London, England, 1934.

3. Communication Networks, Volume II (book), E. A. Guillemin. John Wiley and Sons, Inc., New York, N. Y., 1935, page 162.

No Discussion

Lighting a Network Analyzer Area

V. L. DZWONCZYK
MEMBER AIEE

ELECTRIC power systems have grown tremendously in the first half of the 20th century. This phenomenal increase includes expansion of each of the following components:

1. Greater number of load service points—transmission and distribution stations, industrial plants, and adjacent system interconnections.
2. Greater number of transmission lines, wider range of transmission voltage, and a trend toward higher voltage levels.
3. Greater number of generating points both company owned and those resulting from interconnections.

As the number of loads, paths for power flow, and generator sources increase, the problem of making electrical system calculations of all varieties becomes more complex. To perform these necessary calculations with the aid of only a slide rule would entail a tedious and uneconomical struggle. For this principal reason public utility system planning engineers have made use of more helpful tools known as d-c and a-c network analyzers. The American Gas and Electric Service Corporation, like many other organizations, has hired the use of such analyzers for many years, and for appropriate problems utilized its own d-c analyzer located in the New York office. In June 1951, a modern a-c network analyzer, manufactured by the General Electric Company, similar to one described by Morgan, Rothe, and Winsness[1] was installed in the New York office. Incorporated in this installation is a lighting system which, through its flexibility and character, provides adequate visual conditions for the critical seeing tasks on the horizontal and vertical working planes in the area.

It is the purpose of this paper to analyze the visual requirements in a network analyzer area and to suggest ways and means of providing visual comfort through lighting. Before the lighting facilities are discussed it would appear advisable to review the characteristics of the network analyzer, the nature of work involved in an area housing it, and the seeing tasks which are encountered.

An a-c network analyzer consists of the following fundamental elements:

1. Resistance, inductance, and capacitance units which can be assembled into any desired network to simulate the power systems under study.
2. Adjustable Selsyn generator units operating at 480 cycles per second which can be inserted into the network.
3. Convenient means for interconnecting the various elements.
4. Convenient means of switching indicating meters into any desired part of the network to read voltage, current, reactive, and true power flow.

Figure 1 illustrates the a-c network analyzer for which the lighting to be described was designed.

Description of Analyzers and Work Areas

The work area of the network analyzer group includes rooms where the equipment is located as well as office space where preliminary calculations and final analyses of the analyzer study are made, data are assembled, system one-line diagrams for the problem are prepared, and where conferences may be held. Figure 3 is a plan of the office areas. It will be observed that the d-c and a-c analyzers are located in separate rooms and that the engineers' and conference office also is a separately partitioned space. The a-c analyzer area is provided with a large conference table for more inclusive round-table discussions. All work areas are air-conditioned by means of window units.

The d-c analyzer is composed of 180 adjustable elements assembled within an enclosure approximately $10^1/_2$ by 6 by $6^1/_2$ feet (length, width, height) at the center of which are located the plug and jack connection cords. The indicating instrument is located remotely from the analyzer on the horizontal plane at position x in Figure 3. Beside it is a table upon which work sheets may be placed. The operators face the black-faced analyzer with their backs towards the windows.

The a-c network analyzer illustrated in Figure 1 is composed of 282 adjustable elements and miscellaneous meters mounted on a U-shaped cubicle the main section of which is approximately 26 by 2 feet with two legs each approximately 12 by 2 feet. The board is 7 feet high. Meters are connected into the circuits by push-button control located on the horizontal surface of the operator's desk. In this same vicinity, but on the vertical surface, the light beam galvanometers which indicate the measured quantities are located. The board is predominantly gray in color, having a reflection factor of approximately 40 per cent. The small adjusting knobs are black plastics as are the instrument frames and main control knobs. Except for the latter, directly in front of the operator, the contrasts on the board surface are fairly good, including the operator's push-button panel which is made of gray steel with red plastic buttons.

Visual Tasks

In the office and conference areas the work includes slide-rule calculations, report writing, system map preparation, tabulation of board data, and interpretation of results. This involves visual tasks of a critical nature calling for discrimination of fine detail over prolonged periods of time. The work plane is predominantly horizontal.

Paper 52-119, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE North Eastern District Meeting, Binghamton, N. Y., April 30–May 2, 1952. Manuscript submitted August 31, 1951; made available for printing February 27, 1952.

V. L. Dzwonczyk is with the American Gas and Electric Service Corporation, New York, N. Y.

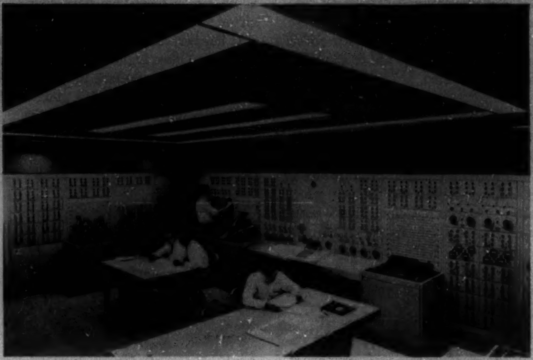


Figure 1. A-c network analyzer—all lights energized

In the a-c analyzer room operations are somewhat as follows:

1. The system network is adjusted on the board with one operator standing within arm's length of it. An associate sitting at a table *Y*, Figure 3, will dictate aloud the values to be adjusted and the connections to be made. In this process critical seeing tasks on both the vertical and horizontal planes are encountered, but each operator can position himself as close to the task plane as is necessary.

2. After the system conditions are established on the analyzer, the required data are read from light beam galvanometers by an operator seated at position *Z* Figure 3; see also Figure 2. The ability to observe these indications is a function of the brightness contrast between the light beam and the translucent background through which it is observed. This condition limits the vertical illumination level in the area. While an operator at the control desk reads aloud values indicated by the meters, his associate records data on appropriate diagrams. Thus the visual tasks are different. A critical task exists on the horizontal plane requiring a relatively high illumination level but the task on the vertical plane must not be illuminated excessively.

A review of these factors suggests a minimum illumination of about 50 foot-candles on the horizontal plane of all areas provided the brightness contrasts also are adjusted satisfactorily. On the vertical plane more than one level of illumination seems desirable to meet the two kinds of critical seeing tasks previously mentioned. It was considered that 40 foot-candles, 66 inches above the floor, should be very satisfactory and that no more than 20 foot-candles, 48 inches above the floor, should be necessary while the galvanometers are being read.

Illumination and Engineering Fundamentals

Having established the character of work and seeing tasks involved, one may now consider the following major fundamentals which must be obtained to effect a reasonable lighting solution:

1. Vertical illumination on the network analyzers should be adequate to allow comfortable vision of the many adjustable resistance, inductance, and capacitance elements and their associated numerical scales, as well as those within the glass-faced instruments.

2. Vertical illumination on these surfaces should not be so high as to cause difficulty in reading the light beam galvanometers.

3. Illumination on the vertical plane should be reasonably diffused to minimize shadows, especially on instrument scales.

4. Light sources must not present themselves as images of specular or veiling glare in glass faces of instruments or other surfaces from operators' usual working positions.

5. Light sources must not present themselves as brightness areas in competition with the object being observed, namely, instrument scales and the markings of adjustable elements on the analyzers.

6. Horizontal illumination on working surfaces should be adequate for the task. In all cases it should be reasonably diffused and the source or any secondary reflecting surface should not cause reflected glare from the horizontal plane.

7. The brightness contrasts between various surfaces in the operator's normal fields of view should not cause the visual surrounding to be uncomfortable.

8. The degree of perfection obtained in each category should justify the actual cost of each incremental gain.

In addition to these lighting fundamentals certain factors which influence lighting must not be disregarded. A good lighting installation requires not only a proper level of illumination, proper quality of illumination, and proper arrangement of lighting fixtures, but also proper reflecting surfaces within the room, both those which make up permanent structures of the building and furniture which is utilized within the area.

Certain factors particularly influence the environment as follows.

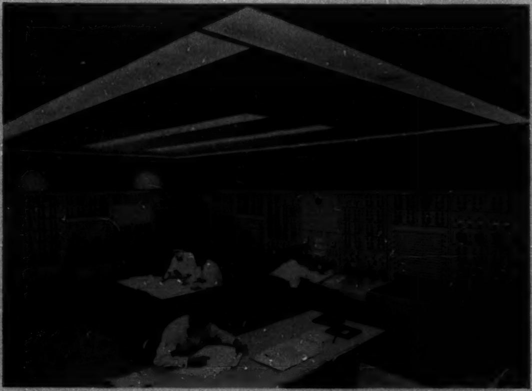


Figure 2. A-c network analyzer—two lamps per troffer adjacent to board de-energized. Same exposure as in Figure 1

Analyzer Surface Finish

In the past, analyzers were predominantly black as in the case of the older d-c board. Gradually through the dissemination of knowledge on brightness contrasts, these backgrounds were changed so that today the choice of pastel shades is accepted. These surfaces should be a matte finish and the color and reflection factor so chosen that visual surrounding is enhanced.

Furniture Finishes

Another important surface is the top of workers' desks. To decrease the ratio of brightness between the work and desk top, it is preferable that desks be selected which have higher reflection factors than common mahogany and dark green linoleum top units. To reduce reflected glare, the top surface should be a matte rather than glossy finish. The use of glass tops should be prohibited.

Room Finishes

The selection of room finishes influences brightness patterns in the office and the relation between brightnesses of walls, ceilings, desks, and the task which is before a worker. Ceiling finish is the most important of factors and it should be selected with the highest reflectance that can be obtained reasonably and should be a matte surface. Ceilings, walls, and floors act as secondary light sources by virtue of their reflecting ability.

It has been recommended by the Committee on Office Lighting of the Illumination Engineering Society that the following reflectances would give good brightness ratios in an office environment:

Ceilings, 85 per cent
Walls, 60 per cent
Desk Tops, 35 per cent
Furniture, 35 per cent
Floors, 30 per cent

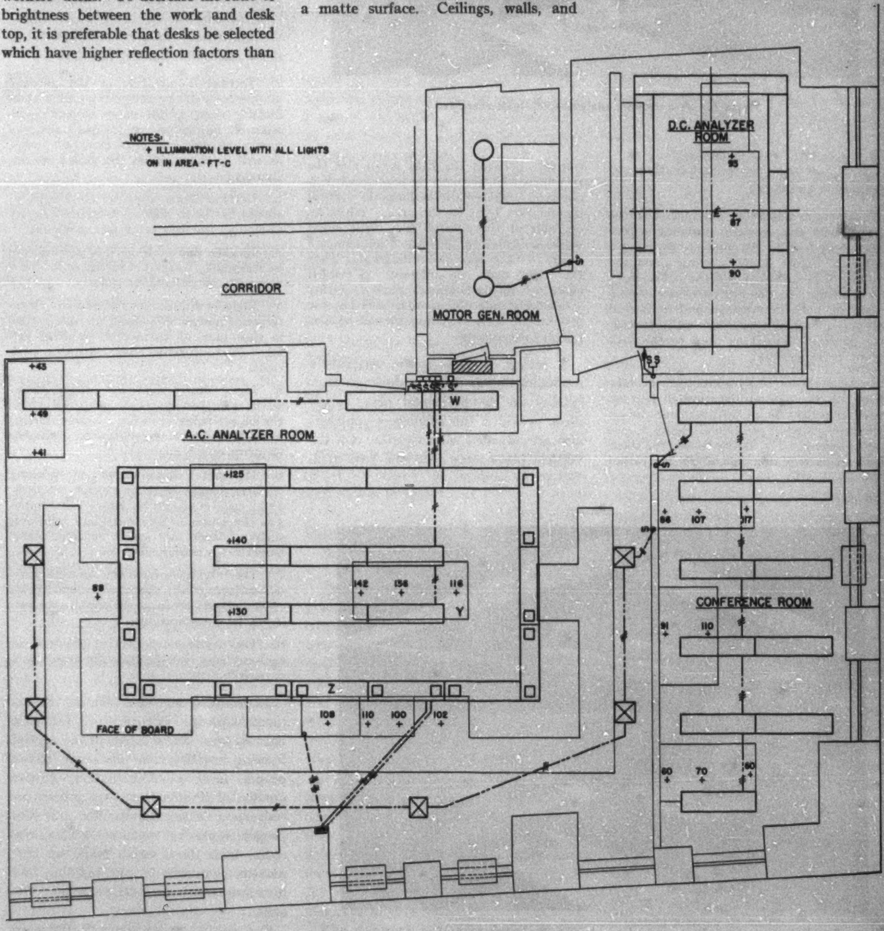

NOTES:
+ ILLUMINATION LEVEL WITH ALL LIGHTS ON IN AREA - FT-C
CORRIDOR
MOTOR GEN. ROOM
D.C. ANALYZER ROOM
95
87
90
SS
W
A.C. ANALYZER ROOM
+43
+49
+41
+125
+140
+130
142
136
116
Y
88
Z
108
110
100
102
FACE OF BOARD
96
107
117
CONFERENCE ROOM
91
110
60
70
60

Figure 3. Plan of network analyzer area

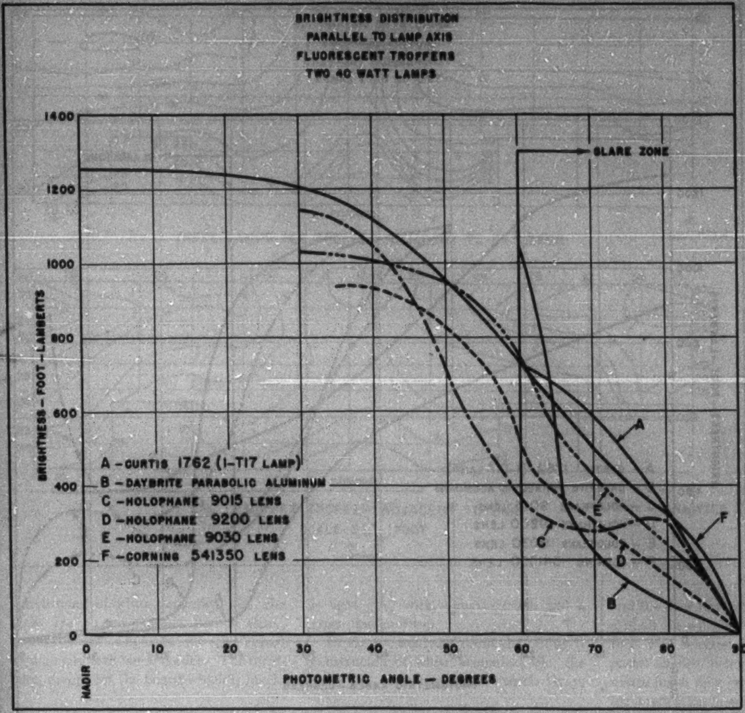


Figure 4. Brightness of troffers parallel to lamps

COLOR

These suggestions permit the use of white or off-white, yellow, ivory, and cream tints, but most often preclude green and blue unless they are not very saturated. It should be noted that acoustical tile and plaster, having sound absorbing cavities, also absorb light in the holes and, therefore, the surface of these tiles should have a high reflection factor and preferably be white in color. The walls should not have reflection factors higher than 60 per cent so that the brightness thereof will not be excessive. It is suggested that, if a dado is used along the lower wall area, a color having a reflectance of at least 30 to 35 per cent is desirable. Similarly, it is beneficial to have floor reflectances as high as 30 per cent and not less than 20 per cent.

Color should be used to make the working environment interesting and pleasant. Cream, ivory, and buff seem warm and they may make an area appear smaller. On the other hand, a room may look larger and cooler by selecting colors such as green, blue green, and blue which seem to recede. Gray finishes are neutral and excellent for either the background or for furniture.

If proper attention is given to the various matters mentioned here, the over-all job will be more nearly correct from the visual standpoint and will supplement and aid the lighting system.

Description of Lighting System and Associated Factors

After considering these factors one must realize also that the selection of a lighting method is controlled to an appreciable extent by aesthetic appearance, ease of maintenance, type of air-conditioning facilities, and atmospheric conditions. The many advantages of fluorescent lighting sources make these lamps the preferred for most office lighting problems, but an a-c source of power is a prerequisite. All conventional methods may find application in lighting an analyzer room and associated areas if the proper determining factors prevail.

For the analyzer area in this instance, it was felt most appropriate to utilize the direct lighting technique using shallow troffers equipped with Holophane *9015* lens plates. The shallow troffers allow a greater unobstructed ceiling height which enhances architectural treatment in the room. Lens plates give the ceiling a neat appearance without such elaboration that the eye is drawn towards it and away from the work. This lens controls brightness in the glare zone very favorably as illustrated in Figures 4 and 5 which show a comparison of several commercially available troffers.

For the most part two lamp troffers equipped for 40-watt hot-cathode fluorescent tubes were specified. In the a-c analyzer room some three lamp units were used in troffers of the rectangle adjacent to the board. These units can have either one, two, or three lamps lighted as desired. Only high-power factor ballasts of a selected group of manufacturers were accepted and watch-dog starters were

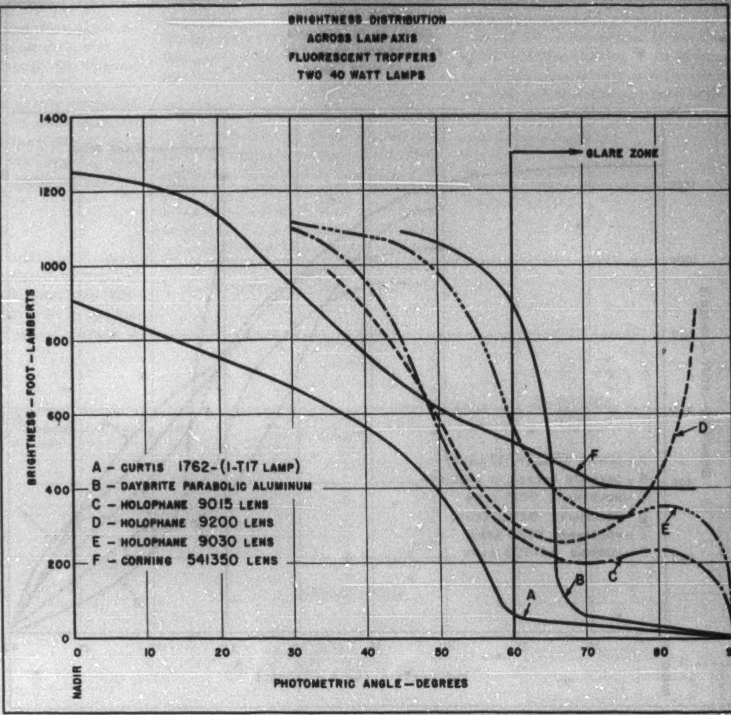


Figure 5. Brightness of troffers perpendicular to lamps

specified. The fixtures were manufactured by Day Brite Lighting Company, Inc., St. Louis, Mo.

Figure 3 illustrates the layout of troffers for the respective rooms in relation to the work areas. Figure 6 illustrates the wiring diagram of the rectangular arrangement of troffers showing the flexibility of lighting that was provided.

The unique conditions of the wiring system include the following:

1. In the d-c analyzer room either the front half, the back half, or the full rectangle may be lighted. When the board is being set up all units lighted are preferable. When readings are being taken, the half away from the board will give a more comfortable situation since all light then would come from behind and the side of operators.

2. In the a-c analyzer room, switching is also flexible. All switches are located at the entrance door, position *W* in Figure 3. Beneath the operator's table at position *Z*, Figure 3, there are located two controls which comprise two groups of 2-way switches the associated units of which are in the ganged-group at the door previously mentioned. Thus at the door one may energize all lighting units and effect control of one, two, or three lamps per troffer adjacent to the analyzer. At position *Z*, an operator also has control of the latter troffers to adjust the lighting level on the analyzer. The effect of having one or three lamps energized in these fixtures can be observed in Figures 2 and 1 respectively. These pictures were taken from the same location with the same exposure.

In the d-c analyzer room the ceiling is

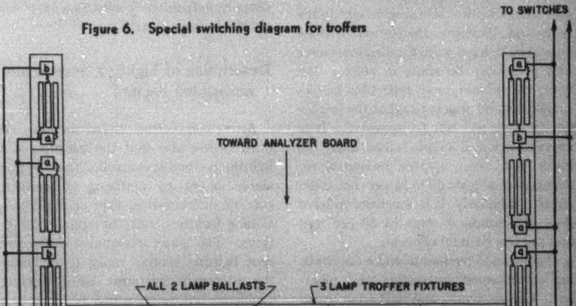


Figure 6. Special switching diagram for troffers

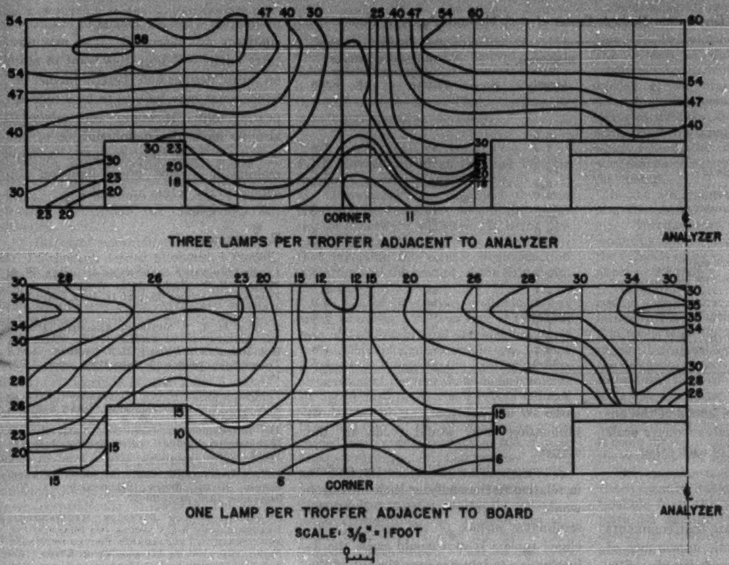


Figure 7. A-c analyzer room—iso-foot-candle chart

constructed of white acoustical tile, the walls are plaster painted French Gray, and the floor has light tan (Greek Skyros) asphalt tile with black border. The furniture, except for the board which is black, is light gray with a matte finish and a green linoleum top.

In the a-c analyzer room the ceiling is constructed of white acoustical tile, the walls are plaster painted French Gray, and the floor has light tan (Greek Skyros) asphalt tile with black border. The furniture was selected in keeping with the pastel shades, being a light gray with a matte finish mist green linoleum top having a reflection factor of 30 per cent. The board is light gray in finish having a reflection factor of approximately 40 per cent.

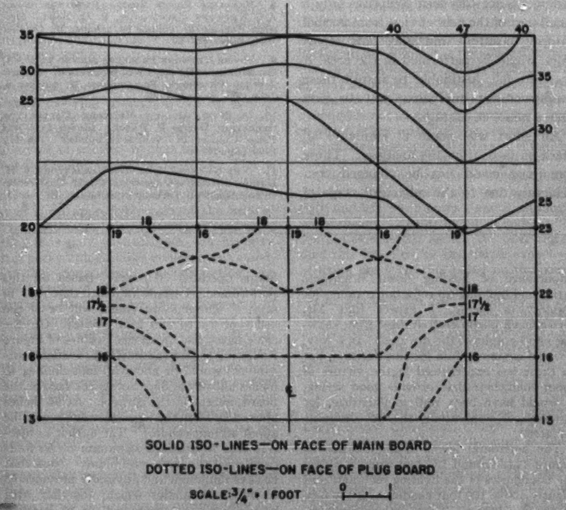


Figure 8. D-c analyzer room—iso-foot-candle chart

The engineer's office also has a white acoustical ceiling, putty gray plaster walls, and Napolean Gray floor tiles with black border. The furniture is similar to that in the other rooms.

Results

The results obtained with this lighting system are illustrated by iso-foot-candle charts, Figures 7 and 8, for the a-c and d-c analyzers respectively. Illumination values on specific horizontal working surfaces are noted on Figure 3. All readings were taken approximately 3 months after the troffers were first energized with no cleaning performed either at the time of installation or in the subsequent interval. It is estimated that the lamps had burned about 600 hours.

Brightness distribution through the center of the a-c network analyzer room at the location of the galvanometers is given in Table I. The maximum brightness ratio is 28.4 and this occurs be-

Table I. Brightness Distribution Through Center of A-C Network Analyzer Room

Surface	Brightness, Foot Length,	Brightness Ratio of Adjacent Surfaces Surface	Ratio
All lights energized			
Acoustical tile above analyzer (a)	6.0		
Back wall (b)	4.3	ab	1.4
Middle of analyzer (c)	21.0	bc	4.9
White galvanometer scale (d)	68.0	cd	3.2
Black control knob (e)	2.4	de	28.4
Gray push-button panel (f)	29.0		
Metal inclined operator's table (g)	20.0	fg	1.5
Floor tile (h)	30.0	gh	1.5
Mist green table top (i)	53.0	hi	1.8
White paper on table (j)	98.0	ij	1.9
U-troffers adjacent to analyzer out			
Acoustical tile above analyzer (a)	6.0		
Back wall (b)	2.4	ab	2.5
Middle of analyzer (c)	7.3	bc	3.0
White galvanometer scale (d)	21.0	cd	2.9
Black control knob (e)	1.2	de	17.5
Gray push-button panel (f)	10.0		
Metal inclined operator's table (g)	7.7	fg	1.3
Floor tile (h)	13.0	gh	1.7

tween the black control knobs of the analyzer and the white galvanometer scale which are within 1 foot of each other.

Conclusions

It is gratifying to state that comments pertaining to the esthetic appearance of the area have been very favorable. The visual feeling one gets as he works in these areas is that of efficiency, neatness, and cheerfulness. The contribution of the lighting system to this atmosphere was obtained at an expenditure for fixtures of $1.39 per square foot.

Although in general it is felt the lighting result is good, two minor conditions call for an explanation. It will be observed from the iso-foot-candle charts of the a-c network analyzer that the illumination in the corner of the analyzer falls to one-half of that on the rest of the surface. This results from a compromise which was made in locating the troffer rectangle. If this rectangle were increased in size by one unit in each long side and if this side were brought one foot closer to the analyzer, an increased illumination level would occur in the corner.

However, this would make the troffers in relation to the analyzer such that they would be specularly reflected from the semimatte surface of the operator's inclined table. It also would cause the illumination to increase near the top of the board at the expense of lower points.

The brightness distribution behind the analyzer in relation to that in front of it might be questioned. This could have been improved by installing indirect lighting units on top of the analyzer which would illuminate the rear walls and ceiling. Since this area comprises only a small part of the field of view from normal working positions and since only repair work would be performed in the area, it was not felt justifiable to install these supplementary luminaires just to gain such a minor advantage.

An effort was made to eliminate all black surfaces from the analyzer. Those remaining could not be changed conveniently due to the advanced stage of production when the matter of brightness contrasts was discussed. The manufacturer might well keep this point in mind on future designs and thereby further enhance what may be said to be a nice-appearing unit. It would seem advisable also to have the metal surface of the operator's table less specular, perhaps making it a mist green linoleum top similar to the tables and desks.

References

1. An Improved A-C Network Analyzer, **W. A. Morgan, F. S. Rothe, J. J. Winsness.** *AIEE Transactions*, volume 68, part II, 1949, pages 891–97.

2. The M. I. T. Network Analyzer, Design and Application to Power System Problems, **H. L. Hazen, O. R. Schurig, M. F. Gardner.** *AIEE Transactions*, volume 49, July 1930, pages 1102–14.

3. Use of A-C Network Calculator in Planning and Operating Electric Power Systems, **P. O. Bobo, H. A. Travers, E. E. George.** *Report 328*, International Conference on Large Electric High-Tension Systems (Paris, France), June 24–July 3, 1948.

4. An Analyzer, **C. Concordia, H. A. Peterson.** *General Electric Review* (Schenectady, N. Y.), September 1945, pages 29–37.

5. Power System Stability Requirements Determined by A-C Calculating Board Study, and Operating Performance Proved by Staged Fault Tests, **W. A. Morgan, Byron Evans.** *Report 319*, International Conference on Large Electric High-Tension Systems (Paris, France), June 24–July 3, 1948.

6. A New A-C Network Analyzer, **H. P. Kuehni, R. C. Lorrain.** *Electrical Engineering* (*AIEE Transactions*), volume 57, February 1938, pages 67–73.

7. Auxiliary Equipment Facilitates Network Calculator Operation, **L. B. Le Vesconte.** *Electric Light and Power* (Chicago, Ill.), volume 24, number 10, October 1946, page 50.

8. Handling Power System Problems on an A-C Network Calculator, **P. O. Bobo.** *Electrical Engineering*, volume 69, number 10, October 1950, pages 867–69.

9. Power Network Problems Solved with A-C Calculating Board, **W. W. Parker.** *Electric Journal* (Pittsburgh, Pa.), volume 32, number 6, June 1935, pages 227–29.

10. A New Aid to Network Calculator Operation, **George P. Wilson.** *Electric Light and Power* (Chicago, Ill.), volume 28, number 1, January 1950, page 78.

11. The Network Analyzer Laboratory, **F. H. Howel.** *Proceedings*, American Power Conference, Illinois School of Technology (Chicago, Ill.), 1952.

Discussion

George J. Taylor (Day Brite Lighting, Inc., St. Louis, Mo.): The author in describing the d-c and a-c analyzer areas refers to the importance of surface finishes in terms of reflectance values. These reflection factors are extremely important to the lighting engineer in analyzing seeing conditions. One gathers that the a-c area is a cheerful one in which to work and this is verified by Figures 1 and 2. The same situation apparently does not exist in the d-c area where it is stated that boards are black faced. What a pity it is to have black boards still in use, especially with our present knowledge of the importance of holding down brightness ratios and building up reflectance values for comfort of seeing. I expect that Mr. Dzwonczyk purposely refrained from showing photographs of the d-c area. In a way, I would commend him for this decision, but to those less experienced in the virtues of room conditions that provide good seeing, it would have been well to illustrate, for purposes of comparison, the less desirable d-c area.

The horizontal levels of illumination in Figure 3 are around 130 foot-candles within the boundaries of the lighting rectangle and slightly above 100 foot-candles near the face of the a-c board. These values were recorded after 3 months at which time the lamps operated 66 hours. Based on this performance, I would estimate that the "in service" illumination values will be of the order of 75 to 100 foot-candles. On the same basis, the "in service" values of Figure 7 showing the iso-candle chart (vertical foot-candles) would be about 35 foot-candles 48 inches above the floor across the face of the board, except in the corners. At 66 inches above the floor the average value would be about 40 foot-candles. The author's objective was to attain a minimum of 50 foot-candles on the horizontal plane. Note that he says minimum and advisedly provides 75 to 100 foot-candles which, together with proper surface reflectances, is in keeping with modern practice. Mr. Dzwonczyk also was striving to get 40 foot-candles on the vertical surface of the board at 66 inches above the floor. For ease of reading lighted galvanometers he desires no more than 20 foot-candles 48 inches above the floor. This he obtains (Figure 7) with one lamp per troffer energized. I wholeheartedly concur with the author's objectives in the prescribed foot-candles and congratulate him on the manner in which he achieved the results with the lighting layout shown in Figure 3. Aside from serving a utilitarian purpose, the lighting arrangement is a pleasant one from the standpoint of esthetics.

Of the eight major fundamentals which Mr. Dzwonczyk desired to obtain, I should like to refer specifically to three. Item 5 which deals with light source (luminaire) brightness is very often overlooked or misunderstood in providing quality illumination. Yet, this is of paramount importance in our consideration of comfortable seeing. The results of luminaire brightness are illustrated clearly in Figures 4 and 5. Only those who have a thorough understanding of illuminating engineering will appreciate the true significance of these curves. The author selected a troffer having a brightness value under 1 candle per square inch in the glare zone (60–90 degrees) when viewing the unit lengthwise (parallel to lamps) or crosswise (perpendicular to lamps). The selection of Holophane (low brightness) 9015 lens plates is superior to other troffers shown for the analyzer room in question. However, it should be pointed out that in larger size rooms the control of brightness in the higher zone angles is of extreme importance. Curve *B* shows that the zonal brightness 70–90 degrees is less than 1/2 candle per square inch. In very large rooms it is highly desirable that lowest possible brightness be attained in the 80–90 degree zone, in which case curve *B* shows brightness at less than 1/4 candle per square inch. Mr. Dzwonczyk is to be complimented for the method used in presenting brightness values in curve form as shown in Figures 4 and 5.

Item 7 relates to brightness contrasts which has a direct bearing with regard to item 8 covering economics. In his closing remarks the author acknowledges the fact that the brightness distribution behind the analyzer in relation to that in the front is not in keeping with good practice. This is shown in Figures 1 and 2 as "black bands" above the analyzer board. Actually, the area is an appreciable part of the field of view, although Mr. Dzwonczyk indicates that it is only a small part. In his very last paragraph he takes an admirable stand, in this connection, by advocating the elimination of all black surfaces from the analyzer board. Since this is suggested to reduce brightness contrast on the face of the board, why not do so above the board also? His suggested solution of installing indirect units is a good one and should have been followed to perfect the problem he started out to solve. To install these units would entail a small additional expenditure. Indeed, if this was added to the $1.39 per square foot for lighting fixtures it would hardly be noticeable. Even if they were appreciably higher, it should be pointed out that the cost of lighting in the analyzer room in all probability is substantially less than 1 per cent of the total investment of other equipment in the room. Consequently, the cost of good lighting is insignificantly small when compared with over-all cost and benefits derived. Good lighting is always a good investment because it deals with the human element which in the final analysis has a direct bearing on morale and operating efficiency.

F. H. Howell* (American Gas and Electric Service Corporation, New York, N. Y.): Mr. Dzwonczyk has described the two principal visual tasks involved in the operation of a network analyzer. It is the discusser's opinion that the more critical one is that involving the actual circuit-selection operation and the reading of the light-beam instruments.

The use of light-beam instruments has many advantages such as rate of response, ease of calibration, and the elimination of the parallax inherent in needle-type meters. However, in designing the lighting system, recognition of the fact that the instruments should be read under conditions of reduced illumination is important. When the analyzer is adjusted, the information to be recorded may involve as many as 500 or more separate readings; this task may last as long as 2 hours and represents a severe test of the adequacy of the lighting system. The operator has to refer constantly to a diagram representing the analyzer circuits, select these circuits, read the meters, and announce their values. Thus his task is on both horizontal and vertical planes. It has been my experience with other installations that this procedure sometimes produces very noticeable eyestrain.

A major source of annoyance elsewhere has been reflections from the meters of light from windows and ceiling lamp fixtures. This effect has been eliminated in our installation, where we have over 100 meters scattered over the analyzer.

Some comments have been offered to the effect that the area between the analyzer top and the ceiling appears as a rather dark background. While it is my understanding that this situation is in conflict with modern lighting principles, its effect is minimized by the rather important fact that the area in question is not at all in the operator's field of vision when seated at the control panel. This remark also applies to a lesser extent to the recorder seated at the nearby table.

When the analyzer was installed, the stainless steel panels at the operator's right and left, see Figures 1 and 2, although wire-brushed, contributed undesirable sources of reflected light. We have remedied this situation by the simple expedient of fixing green blotting paper on the panel surfaces.

Our analyzer has been in service for almost a year. During that time it has been in constant use and it is my opinion as an operator that the lighting system is entirely adequate.

W. H. Kahler (Westinghouse Electric Corporation, Cleveland, Ohio): This is a particularly good paper because it illustrates completely some important principles of illuminating engineering. The seeing tasks have been appraised carefully and illumination fundamentals have been applied to produce a successful result.

Considering all requirements and limitations, the direct troffer lighting system appears to be adequate and comfortable. The success of direct lighting can be attributed to careful specification of the finishes of the major reflecting surfaces throughout the room. The high reflection factor of the floor, desk tops, and wall finishes results in illumination on the ceiling which relieves the otherwise harsh brightness difference between the luminaires and the ceiling.

In the a-c analyzer room the dark area over the control panels is slightly distracting. This dark area could be eliminated by mounting indirect luminaires on the top of the cubicals or by extending the face of the panel to the ceiling.

The light beam galvanometer is the most critical seeing task in the entire room. This instrument is easier to read at low illumination levels while all other tasks in the room require high level lighting. In the future, the designers of such instruments should try to improve the visibility by using a larger scale and a brighter light beam.

Other instruments on the analyzer panel also are difficult to read and there certainly is room for improving the instruments from the standpoint of visibility.

The specular metal surface on the inclined operator table has been found objectionable from the standpoint of reflected glare and in future designs this shiny surface should be replaced by a light nonspecular material.

Thus it can be seen that the lighting engineers do not have complete control of the visibility conditions of equipment of this type. The equipment designers should consider visibility of instruments as well as electrical performance to minimize the visual effort required of the operators.

E. A. Lindsay (General Electric Company, Cleveland, Ohio): The analyzer room discussed in this article is a comfortable and efficient work space.

Mr. Dzwonczyk certainly is to be commended for his recognition of the importance of having comfortable brightness relationships to provide easy seeing by eliminating the distracting influence of black instrument frames, black control knobs, and so forth. As a first step, he obtained a few meters with light finishes. Possibly with further encouragement the meter manufacturers will make their equipment as suitable for light finished boards as the present equipment is for the older black boards.

As in nearly all installations there may be details that could be improved. For example: Other areas of black that might be lightened are the surface of the plug board and the feature strip around the bottom of the board. These areas could be very much lighter than they are now, with possibly a color difference where it is desirable to set them off from adjacent surfaces.

It appears from the illustration that the brightness of the area immediately above the board (a portion of the upper sidewalls and ceiling) is considerably lower in brightness than other areas of the surroundings. This is borne out by the brightness values in the text and by actual observation of the room. It would appear that this condition could be improved with some indirect light as mentioned in the article. Possibly a row of *T12* fluorescent lamps on the top of the board would accomplish this.

Maintaining good visibility on the galvanometer scales might be handled without reducing the brightness of the board if small shields were placed over each scale. This should be preferable to reducing the board illumination.

The suggestion in the paper that light linoleum be substituted for the specular metal desk surface adjacent to the area *Z* is an excellent one and should be passed on to the supplier of the analyzer.

Kurt Franck (Holophane Company, Inc., Newark, Ohio): Having had the opportunity of personally inspecting the lighting installation described in Mr. Dzwonczyk's paper, I should like to say that the solution to the problem and the over-all effect of the visual results are excellent and that the following remarks cover only some minor details for Mr. Dzwonczyk's consideration.

If the lighting problem were only one of concentrating as much light as possible on the vertical control boards, it might be well to use concentrating type prismatic lenses so as to reach clear down to the bottom of the board. Actually, as the work being done in the network analyzer area includes quite a number of tasks beyond operation of the control boards themselves, Mr. Dzwonczyk's solution of using general illumination prismatic lenses is probably best all around. However, it does sacrifice uniformity of illumination on the vertical boards by not providing high values of illumination along the bottom line of the boards. This is visually noticeable at the extreme right-hand corners of Figures 1 and 2 where the board reaches down to the floor. Also it is apparent from the iso-foot-candle charts of Figure 7 which show foot-candle values at the bottom of the boards to be only one-third or one-fourth of those at the top of the board.

The arrangement of the lens units in the form of a rectangle as laid out by Mr. Dzwonczyk for the analyzer area somewhat penalizes the illumination values at the corners of the area. It must be realized that there is a greater summation of light, contributed by each side of the rectangle, at the lateral center of each control board wall than at its corners, and that the illumination toward the corners of the area is bound to suffer. The iso-foot-candle charts of Figure 7 plainly show this, indicating that the average illumination at the corners is approximately half of that obtained at the lateral center of each control board. This condition might have been remedied by providing a greater concentration of and lamps at the corners even if it meant deviating somewhat from the clean lines of the rectangular arrangement of luminaires.

Mr. Dzwonczyk calls attention to the fact that the brightness distribution beyond the analyzer in relation to that in front of it is not too good. As both Figures 1 and 2 indicate, there is quite an extended zone of dark area above the analyzer control boards which as Mr. Dzwonczyk states could have been taken care of easily by the use of a few indirect type luminaires in the areas behind the analyzer. Mr. Dzwonczyk considers this a minor point and it certainly is. Yet, in my opinion, it would have been aesthetically well worth while and economically quite reasonable to give a final touch to this outstanding lighting installation by illuminating the extended dark areas behind the control boards of the analyzer area.

H. E. D'Andrade (General Electric Company, New York, N. Y.): The illumination of a network analyzer is literally a new project in the field of lighting. It is true that an analyzer room resembles a control room in certain respects, but in the analyzer room the lighting task differs in the following ways:

1. Many dials have to be set to predetermined readings with a high degree of accuracy, whereas in a control room the dials are not set but are read periodically in much the same way that a meter reader scans a kilowatt-hour meter.

2. The analyzer room has certain scales which must be read with micrometer accuracy. Thus it appears that the seeing task is of a more critical nature than in a control room—using the word critical as applied only to seeing.

There is another type of work done which demands good horizontal illumination—recording readings on diagrams and so forth. From these requirements it appears that no single lighting solution would be satisfactory.

Mr. Dzwonczyk is to be congratulated on his decision to provide two lighting systems: one for vertical illumination and the other for horizontal illumination. In addition, through flexible wiring circuits it is possible to obtain several combinations, and this is perhaps the most commendable part of the design.

Experience with control rooms indicates that operators like a lighting system with versatile controls, enabling an operator to choose the lighting effect which he finds most satisfactory. Mr. Dzwonczyk's system meets these requirements, though it might be added that there is no indirect system in the room. Possibly a small indirect component might add to the comfort and versatility of the present lighting system. This, however, is a minor detail, and since in every other respect the design conforms to the best modern practice, it might almost be hailed as a model.

The seeing problem has been properly analyzed and it has been well solved. The lighting levels are adequate. The brightness ratios are fairly comfortable. It is gratifying to sze a difficult lighting problem solved so satisfactorily.

V. L. Dzwonczyk: It is very gratifying to receive these interesting discussions from such well-informed gentlemen. This paper is made materially more valuable by their informative contributions.

Of the twelve pertinent topics discussed, the brightness contrast between the areas above and behind the analyzer in relation to that in front of it was mentioned by all men. It, therefore, deserves major attention. The specular operator's table top, the blackness of plug boards, instruments, and knobs, and the seeing task at the galvanometer were also discussed more than once.

There are several ways in which the analyzer area could be illuminated and there are many degrees of perfection one could envision. While it is evident from the paper that I appreciate the value of low brightness contrasts and have made provisions to obtain them to a degree, it may not be so clear why the extra money was not spent to purchase indirect lighting fixtures that would illuminate the ceiling and wells above and behind the analyzer to accomplish the following:

1. Increase the brightness of the walls and ceiling behind and above the analyzer to reduce the brightness contrasts between these areas and the analyzer face.

2. Increase the ceiling brightness to lower the contrast between it and the adjacent lighting troffers.

If both of these conditions were satisfied, a more perfect solution would have been obtained. Several questions seem pertinent, however. How great a part does this area comprise of the whole field of view from normal working positions? How much benefit would be realized by the added expenditure as measured by comfort to workers and work output? These are only two questions that could be asked. My firm belief is that lighting should be engineered and thus one should include economics as one factor affecting the solution.

As stated in the paper, the projected visual area in question from the most severe normal seated position in the room is small compared to the total view. It is noted that the discussers vary in the degree of evaluation of this point. Mr. Taylor, who is as good a lighting engineer as I know, feels that it would not have been a unreasonable expenditure to add the indirect component of light. I sincerely respect his judgment, however, to add it; one is not only faced with the purchase and installation of lighting equipment but also with the problem of providing a suitable shield along the top of the analyzer so as to hide the fixtures from view and not destroy the aesthetic appearance of the installation. It isn't that the cost would have added an appreciable percentage to the total outlay for all equipment in the room; but would it have realized an incremental gain commensurate with the expenditure? Mr. Howell, who has had considerable experience with several analyzers and is thoroughly familiar with problems associated with them has this to say, "... the area in question is not at all in the operator's field of vision when seated at the control panel. This remark also applies to a lesser extent to the recorder seated at the nearby table." The discussers' comments regarding this deficiency are good and fundamentally correct; but the desire for perfection should not exclude consideration of the cost.

I am extremely happy to note the supporting comments regarding the black surfaces on the analyzer by Mr. Kahler and Mr. Lindsay. Each of these men, and there are others, believe that the instruments and other accessories placed on analyzers and control boards should be pastel in color rather than black, yet their companies' designers of products other than for lighting requirements do not heed their preachings. How can engineers interested in lighting problems get all instrument manufacturers to recognize that the value of their instrument would be materially enhanced by improving their visibility?

The specular table top is another instance where a considerable improvement could have been made if the designers were sufficiently informed of the simple lighting fundamentals that are involved.

Comments by Mr. Franck on the availability of other prismatic lenses which would have helped to produce a more evenly distributed illumination on the vertical surface of the analyzer are appreciated. As in all lighting problems, compromises must often be made. Our desire to make use of a uniform and low brightness lens influenced the choice. I wonder, however, if the lenses to which Mr. Franck referred would accommodate the use of more than one lamp per troffer as is necessary here?

Mr. D'Andrade's comment on the flexibility of lighting systems is good and one with which we agree perfectly. It is one way to meet conflicting requirements as encountered in this problem and also to solve the classic fact that you cannot satisfy all of the people all of the time.

Semifinite Terminated Electric Wave Filters

SHELDON S. L. CHANG
ASSOCIATE MEMBER AIEE

Synopsis: Specially designed filters to work between a finite resistance at one end and an open circuit or a short circuit at the other end can be made to duplicate exactly the attenuation and phase-shift characteristics of image parameter filters terminated in equal resistances. By doing so a 6-decibel flat gain and considerably higher frequency limit due to parasitic elements can be obtained without added cost. The design procedure follows exactly that of the well-known image parameter filters. A set of algebraic equations express the new element values in terms of the values for image parameter filters.

Semifinite terminated filters also can be designed on the insertion loss basis. The assumption of matching impedance to one terminal only simplifies the process of computation considerably, and curves for direct reading of element values are prepared for the 2-section filters to meet moderate discrimination requirements.

While the assumption of open circuit or short circuit cannot be strictly realized, these filters are applicable wherever the ratio of terminating resistances exceeds a certain limit. This minimum ratio of termination as a function of the discrimination requirements and the allowable passband distortion is calculated and plotted into curves. Its value is between 5 and 50 for most applications.

MOST filters are designed to be terminated with equal resistances, except in band-pass filters where impedance transformation sometimes is applied. When the termination at one end is a very high impedance, such as the grid circuit of a class-*A* amplifier, the usual practice is to shunt the high-impedance end with an equal resistance, and to design the filter as one to be terminated with equal resistances. This is not necessarily the best arrangement. The present study on semifinite terminated filters is aimed at improvement of the performance of filters with such terminations by designing them directly for the open-circuit termination at one end. It is shown here that with the same number of elements as usually used, the latter method gives a receiving-end voltage twice as great at all frequencies. Besides, the modified element values make it possible to absorb about 60 per cent more capacitance at the receiving end if the filter is a low-pass or band-pass filter.

In other words, up to a 6-decibel gain and a higher limit of cutoff frequency is obtainable without added cost. This method is applicable to low-pass, high-pass, band-pass, and band-elimination filters. In the case of band-pass filters, impedance transformation can be introduced to increase further the receiving-end voltage if so desired.

This method of designing filters with very high terminating impedances at the receiving end can be generalized to permit the design of filters with very low terminating impedances at the receiving end. In such cases the receiving-end current is almost independent of the receiving-end impedance, and the attenuation and phase shift can be defined conveniently in terms of the current ratio. The method also can be generalized to include filters terminated with very low or with very high impedances at the sending end and with a constant resistance at the receiving end.

Ordinary filters are applicable only when the two terminating resistances are of prescribed values. The semifinite filters can be used where the terminating resistances are widely different. The minimum ratio of terminating resistances is found to be a function of the allowable passband frequency distortion and the sharpness of rise of the attenuation curve. Values are calculated and plotted. The minimum ratio is well within practicable limits.

Besides the natural applications of the semifinite filters, it is sometimes advantageous to use a synthetic low-impedance termination. Channel filters with parallel connection at one end are examples of such an application. In ordinary cases, the design of each filter is complicated by the shunt effect of the driving-point impedances of all the other paralleled filters. The use of semifinite terminated filters with a shunt feedback amplifier at the receiving end in place of the low-impedance termination allows each filter to be designed individually.

Types and Configuration of Semifinite Terminated Electric Wave Filters

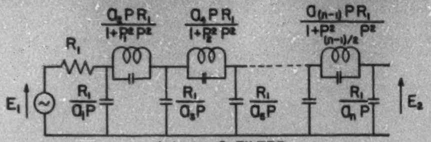
Semifinite terminated electric wave filters are defined as filters terminated with a finite resistance at one end, and a short circuit or an open circuit at the other end. They can be classified either according to their frequency range or according to their terminations. Classified according to their frequency range, there are the low-pass, high-pass, band-pass, and band-elimination filters. The constants of the three last types of filters can be derived from that of the low-pass filters with a suitable frequency transformation.

However, a more direct method is to let the frequency variable *P* represent different functions of frequency for different filters as given here:

Paper 52-123, recommended by the AIEE Special Communications Applications Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE North Eastern District Meeting, Binghamton, N. Y., April 30–May 2, 1952. Manuscript submitted September 10, 1951; made available for printing February 28, 1952.

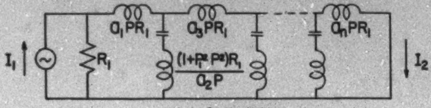
Sheldon S. L. Chang is with Robbins and Myers, Inc., Springfield, Ohio.

This paper represents the essential material of a thesis accepted in 1947 by the Faculty of Purdue University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. This work was suggested and supervised by Prof. R. P. Siskind whose generous guidance and encouragement are gratefully acknowledged.

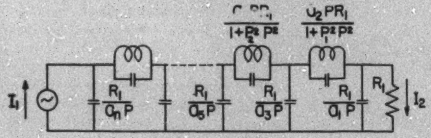


$\frac{a_1 P R_1}{1+P_1^2 P^2}$
$\frac{a_4 P R_1}{1+P_2^2 P^2}$
$\frac{a_{(n-1)} P R_1}{1+P_{(n-1)/2}^2 P^2}$
R_1
E_1
$\frac{R_1}{a_1 P}$
$\frac{R_1}{a_3 P}$
$\frac{R_1}{a_5 P}$
$\frac{R_1}{a_n P}$
E_2

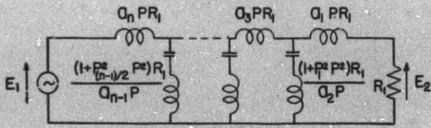
(a) 1-∞ FILTER



(b) 1-0 FILTER



(c) ∞-1 FILTER



$a_n P R_1$
$a_3 P R_1$
$a_1 P R_1$
E_1
$\frac{(1+P_{(n-1)/2}^2 P^2) R_1}{a_{n-1} P}$
$\frac{(1+P_1^2 P^2) R_1}{a_2 P}$
R_1
E_2

(d) 0-1 FILTER

Figure 1. Configurations of semifinite terminated filters

Low-pass filters: $P=j\frac{f}{f_c}$

High-pass filters: $P=-j\frac{f_c}{f}$

Band-pass filters: $P=j\dfrac{\frac{f}{f_m}-\frac{f_m}{f}}{\frac{f_2}{f_m}-\frac{f_m}{f_2}}$ (1)

Band-elimination filters: $P=-j\dfrac{\frac{f_2}{f_m}-\frac{f_m}{f_2}}{\frac{f}{f_m}-\frac{f_m}{f}}$

where f_c is the cutoff frequency of a low-pass or a high-pass filter, f_1 and f_2 are the lower and higher cutoff frequencies respectively of a band-pass or a band-elimination filter, and f_m is the geometrical mean $\sqrt{f_1 f_2}$.

These expressions for P determine the nature of the filter elements. The same notation aP for a reactance element may represent a coil, a capacitor, a series-resonant circuit, or a parallel-resonant circuit in the cases of low-pass, high-pass, band-pass, or band-elimination filters respectively.

The discrimination requirements in general can be expressed in terms of P instead of frequency. That makes it possible to convert a low-pass filter into any of the three other types to meet the specifications. For convenience in drawing, the configurations of low-pass filters are given in all illustrations.

Classified according to their terminations, the four types of semifinite filters are:

1. One inserted between a source with finite resistance and an open circuit, which will be referred to as a 1-∞ filter.
2. One inserted between a source with finite resistance and a short circuit, which will be referred to as a 1-0 filter.
3. One inserted between a constant current source and a finite resistance, which will be referred to as a ∞-1 filter.
4. One inserted between a constant voltage source and a finite resistance, which will be referred to as a 0-1 filter.

The insertion ratio $\epsilon^{\alpha+j\beta}$ is defined as the ratio of voltages or currents at the receiving end without and with the filter. For 1-∞ and 0-1 filters, it is defined as the voltage ratio. For 1-0 and ∞-1 filters, it is defined as the current ratio. α is the attenuation or insertion loss in nepers, and β is the phase shift in radians.

Configurations of semifinite, low-pass filters are given in Figure 1. The insertion ratios are readily verified to be

$$\begin{aligned} &\text{1-}\infty\text{ filter:} && \epsilon^{\alpha+j\beta}=\frac{E_1}{E_2}\\ &\text{1-0 filter:} && \epsilon^{\alpha+j\beta}=\frac{I_1}{I_2}\\ &\infty\text{-1 filter:} && \epsilon^{\alpha+j\beta}=\frac{I_1}{I_2}\\ &\text{0-1 filter:} && \epsilon^{\alpha+j\beta}=\frac{E_1}{E_2} \end{aligned} \qquad (2)$$

The coefficients a_k give the values of reactances of the series arms and the values of susceptances of the shunt arms.

The nth elements of 1-∞, and ∞-1 filters are always shunt arms, and the nth elements of 1-0 and 0-1 filters are always series arms. An impedance in series with an open circuit or in parallel with a short circuit is useless. The first element may be a series or a shunt arm, depending on the number of elements being odd or even.

With the numbering system previously stated, $\epsilon^{\alpha+j\beta}$ will be the same for all four types of filters if the corresponding a_k's are the same. The equivalence between 1-∞ and 1-0 filters is obvious because they are inverse forms of each other. The equivalence between 1-∞ and ∞-1 filters is established by the reciprocity theorem in the following manner: In Figure 2 suppose I is the current induced at the receiving end by the voltage E applied at the sending end and I' is the current induced at the sending end by the voltage E' applied at the receiving end. Then from the reciprocity theorem

$$I/E=I'/E'$$

Multiplying both sides by R_∞ this equation can be rewritten as

$$IR_\infty/E=I'/(E'/R_\infty)$$

Now let R_∞ and E' approach infinity in such a way that E'/R_∞ is a constant. Then the left-hand termination of the filter becomes either an open circuit with IR_∞ as the open-circuit voltage or a con-

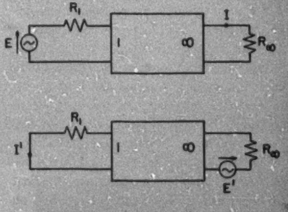


Figure 2. Diagrams showing the equivalence between 1-∞ and ∞-1 filters

stant-current generator with E'/R_∞ as the generated current. IR_∞/E becomes the voltage ratio or $\epsilon^{-\alpha-j\beta}$ of a 1-∞ filter, while $I'/(E'/R_\infty)$ becomes the current ratio or $\epsilon^{-\alpha-j\beta}$ of a ∞-1 filter. Thus the equivalence between these two types of filter has been proved.

There are a few direct applications of semifinite terminated filters: A 1-∞ filter can be used as an input filter to a vacuum-tube amplifier. A ∞-1 filter can be used in the plate circuit of a pentode tube or a series feedback amplifier. A 1-0 filter can be used as an input filter to a shunt feedback amplifier. A 0-1 filter can be used as an output filter of a shunt feedback amplifier.

Equivalent Image Parameter Filters

An equivalent image parameter filter is defined as one which, when terminated with finite resistance at one end and an open circuit or a short circuit at the other end, gives the same insertion loss and phase shift as does an image parameter filter terminated with equal resistances. The equivalence is established by equating the algebraic expressions of the two insertion ratios.

For a filter whose configuration is given in Figure 1, the number of coefficients of the insertion ratio expression is exactly equal to the number of elements. Thus for one image parameter filter there is one and only one corresponding equivalent image parameter filter.

To illustrate the method, the equivalent image parameter filter of 1-section prototype filter is derived now. The configuration of a mid-shunt prototype filter is given in Figure 3(A). Since $L_k=R_1/\pi f_c$ and $C_k=1/\pi f_c R_1$, the reactances represented by L_k and C_k are respectively

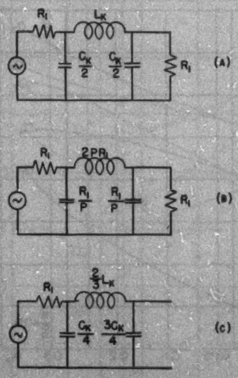
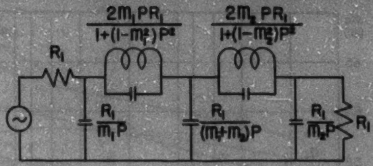


Figure 3. A 1-section prototype filter and its equivalent semifinite filter

Figure 4. A 2-section m-derived filter with its element values expressed in terms of P



$$j\omega L_{kk}=2j\frac{f}{f_c}R_1=2PR_1 \qquad (3)$$

and

$$\frac{1}{j\omega C_k}=\frac{f_cR_1}{2jf}=\frac{R_1}{2P} \qquad (4)$$

Hence the element values can be written in terms of P as given in Figure 3(B). The voltage ratio is readily shown to be

$$E_1/E_2'=2+4P+4P^2+2P^3 \qquad (5)$$

$$\epsilon^{\alpha+j\beta}=E_1/2E_2'=1+2P+2P^2+P^3 \qquad (6)$$

For the equivalent filter of a prototype filter, the constants $P_1, P_2, \ldots$ in Figure 1 are identically zero, and the shunt capacitances in the series arms, as well as the series inductances in the shunt arms, are removed from the diagrams. For $n=3$, the insertion ratio is

$$\epsilon^{\alpha+j\beta}=E_1/E_2=1+(a_1+a_3)P+a_2a_3P^2+a_1a_2a_3P^3 \qquad (7)$$

Equating 6 and 7 yields

$$a_1+a_3=2, \qquad a_2a_3=2, \qquad a_1a_2a_3=1$$

Solving these equations gives

$$a_1=1/2, \qquad a_2=4/3, \qquad a_3=3/2$$

If the element values are written in terms of L_k, C_k, this filter becomes the one shown in Figure 3(C).

By a similar process, the equivalent prototype filters with different number of elements may be derived. The element values are listed in Table I.

Since prototype filters of more than three sections are rarely used, Table I should be sufficient. However, it can be proved by mathematical induction that with the exception of the last element each coefficient a_k is larger than a_{k-1} and that $a_{k\neq n}$ approaches 2 as its ultimate limit. The coefficient of the last element, a_n, is also monotonically increasing and its ultimate limit lies between 1.70690 and 1.70712.

In general, let three sequences of numbers t_k, b_k, and c_k be defined by the following equations

$$t_2=2 \qquad t_3=9,$$

$$t_k=\frac{(1+t_{k-1})^2}{t_{k-2}} \qquad K=4, 5, 6 \ldots \qquad (8)$$

$$b_1=\frac{1}{2},\ b_k=\frac{2t_k}{1+t_k} \qquad K=2, 3, 4 \ldots \qquad (9)$$

$$c_2=1,\ c_k=\frac{2C_{k-1}}{b_{k-1}} \qquad K=3, 4, 5 \ldots \qquad (10)$$

The value of the kth element in an equivalent prototype filter of n elements is

$$\begin{aligned} a_k&=b_k && \text{for } k=1, 2 \ldots n-1\\ a_k&=c_k && \text{for } k=n \end{aligned} \qquad (11)$$

The derivation of the preceding relations is lengthy and will not be included in this paper.

To illustrate the method used in determining the element values of equivalent m-derived filters, the 2-section case is taken as an example. An ordinary 2-section m-derived filter with its element values given in terms of P is shown in Figure 4. Its insertion ratio is easily shown to be

$$\epsilon^{\alpha+j\beta}=1+A_1P+\frac{A_2P^2+A_4P^3}{1+(1-m_1^2)P^2}+\frac{A_3P^2+A_5P^3}{1+(1-m_2^2)P^2} \qquad (12)$$

where

$$A_1=2(m_1+m_2) \qquad (13)$$

Table I. Element Values of Equivalent Prototype Filters

Number of Elements	a_1	a_2	a_3	a_4	a_5	a_6	a_7
2	1/2	1					
3	1/2	4/3	3/2				
4	1/2	4/3	9/5	5/3			
5	1/2	4/3	9/5	100/51	17/10		
6	1/2	4/3	9/5	100/51	289/145	29/17	
7	1/2	4/3	9/5	100/51	289/145	3,364/1,683	99/58

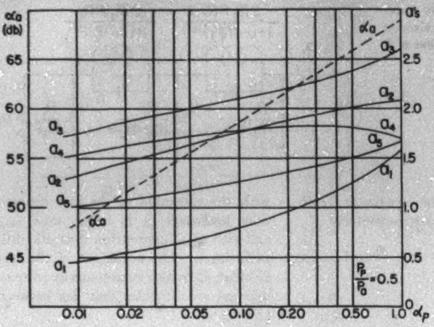


Figure 5. Element values of semifinite terminated filters designed on insertion loss basis. k = 0.500, $P_1=0.436$, $P_2=0.677$

$$A_2=2\frac{m_1^2(m_1+m_2)}{m_1-m_2} \tag{14}$$

$$A_3=-2\frac{m_2^2(m_1+m_2)}{m_1-m_2} \tag{15}$$

$$A_4=m_1\left(\frac{m_1+m_2}{m_1-m_2}\right)(2m_1^2-1) \tag{16}$$

$$A_5=m_2\left(\frac{m_1+m_2}{m_1-m_2}\right)(1-2m_2^2) \tag{17}$$

The insertion ratio of a 2-section semifinite filter can be expressed in the same way as equation 12, with the coefficients A_k given in the following ($P_1^2=1-M_1^2$, $P_2^2=1-M_2^2$)

$$A_1=a_1+a_3+a_5 \tag{18}$$

$$A_2=a_2a_3+a_2a_5+\frac{a_2a_3a_4a_5}{m_1^2-m_2^2} \tag{19}$$

$$A_3=a_4a_5-\frac{a_2a_3a_4a_5}{m_1^2-m_2^2} \tag{20}$$

$$A_4=a_1A_2 \tag{21}$$

$$A_5=a_3a_4a_5+a_1A_3 \tag{22}$$

Solving a_1 from equations 21, 14, and 16

$$a_1=\frac{2m_1^2-1}{2m_1} \tag{23}$$

From equations 22, 15, and 17

$$a_3a_4a_5=\frac{m_2(m_1+m_2)}{m_1}(1+2m_1m_2) \tag{24}$$

From equations 18 and 23

$$a_3+a_5=\frac{1}{2m_1}(2m_1^2+4m_1m_2+1) \tag{25}$$

Substituting equations 25 and 24 in equation 19 and solving for a_2, there results

$$a_2=\frac{4m_1^3}{2m_1^2+1} \tag{26}$$

From equations 20, 24, and 26

$$a_4a_5=\frac{2m_2(m_2+2m_1+2m_1^2m_2)}{1+2m_1^2} \tag{27}$$

Dividing equation 24 by equation 27

$$a_3=\frac{(m_1+m_2)(1+2m_1m_2)(1+2m_1^2)}{2m_1(m_2+2m_1+2m_1^2m_2)} \tag{28}$$

From equations 28 and 25

$$a_5=\frac{(1+2m_1m_2)^2+2(m_1+m_2)^2}{2(2m_1+m_2+2m_1^2m_2)} \tag{29}$$

Dividing equations 27 and 29

$$a_4=\frac{4m_2(2m_1^2m_2+2m_1+m_2)}{(1+2m_1^2)[(1+2m_1m_2)^2+2(m_1+m_2)^2]} \tag{30}$$

For a_1 to be positive, m_1 must be larger than $1\sqrt{2}$. Since m_1 and m_2 can be interchanged without varying the insertion ratio, this requires that one of the m values be larger than 0.707.

By a similar process, the element values of a $1^1/_2$ section equivalent m-derived filter is calculated to be

$$a_1=\frac{1}{2} \tag{31}$$

$$a_2=\frac{2m(1+m)}{1+2m} \tag{32}$$

$$a_3=\frac{m(1+2m)^2}{1+2m+2m^2} \tag{33}$$

$$a_4=\frac{1+2m+2m^2}{1+2m} \tag{34}$$

In the foregoing analysis, a method has been described by which the element values of the equivalent image parameter filters can be calculated. However, all the cases shown have prototype impedances in the original filter. In the case

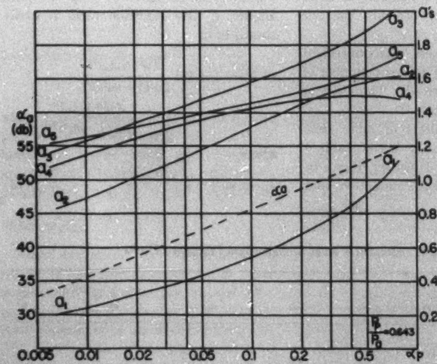


Figure 6. Element values of semifinite terminated filters designed on insertion loss basis. k = 0.643, $P_1=0.512$, $P_2=0.771$

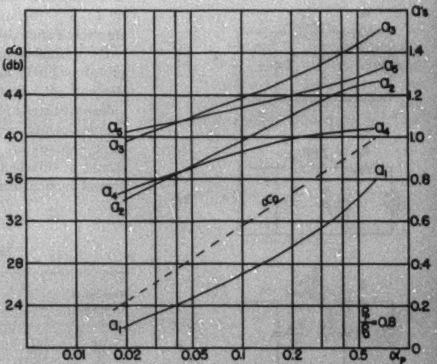


Figure 7. Element values of semifinite terminated filters designed on insertion loss basis. k = 0.800, $P_1=0.610$, $P_2=0.868$

where the original filter has m-derived image impedance, the equivalent filter contains negative elements. Thus the application of the equivalent image parameter filters is limited to cases with not too critical passband requirements. When low passband distortion is required, it is necessary to use the insertion loss method of design.

Semifinite Terminated Filters Designed on Insertion Loss Basis

The insertion loss method of designing semifinite terminated filters is included for completeness and for comparison, and also to present the curves giving direct design information. Although the method used to calculate the element values is different, the same results can be obtained by following Darlington's[2] original work.

Instead of using single sections of image parameter filters as building blocks, the insertion loss method as introduced by Norton[1] and Darlington[2] assumes as a starting point the power ratio for a ladder-type filter of l sections

$$\epsilon^{2\alpha}=1-\left[\frac{SP(Q_1^2+P^2)(Q_2^2+P^2)\ldots(Q_l^2+P^2)}{(1+P_1^2P^2)(1+P_2^2P^2)\ldots(1+P_l^2P^2)}\right]^2 \tag{35}$$

where $Q_1, Q_2 \ldots$ and $P_1, P_2 \ldots$ are real constants less than unity, and S is an arbitrary constant.

Equation 35 implies that there are a number of frequencies in the passband for which the insertion loss α is zero, and a number of frequencies in the attenuation band for which the insertion loss is infinite. These frequencies satisfy the equations $P=\pm jQ_1$, $P=\pm jQ_2 \ldots$ and $P=\pm j/P_1$, $P=\pm j/P_2 \ldots$ respectively. Thus the power ratio as given in equation 35 is not in its most general form but is restricted to that of a group of most useful filters.

The choice of the constant S depends on the amount of allowable ripple in the passband. From equation 35 it is apparent that doubling the value of S would double the passband ripple and raise the insertion loss in the attenuation band by 3 decibels, while leaving the general shape of the attenuation curve unaltered.

One notes that the power ratio is equal to the product of the insertion ratio with its complex conjugate function along the real frequency axis that is, the imaginary axis on the P plane. The insertion ratio cannot be uniquely determined from the power ratio without additional restrictions. It is well known that any impedance or transfer function of a passive physical network cannot have roots and zeros of P with positive real components, and this condition is exactly what one needs to determine the insertion ratio uniquely. Since the power ratio is a real function of P^2, its zeros are symmetrically located in the four quadrants of the P plane. Only the zeros with negative real components are selected to form the insertion ratio. Since these zeros are in conjugate pairs, the resulting insertion ratio can be written as

$$\epsilon^{\alpha+j\beta}=\frac{(P+e_0)(P^2+h_1P+g_1)(P^2+h_2P+g_2)\ldots(P^2+h_lP+g_l)}{(1+P_1^2P^2)(1+P_2^2P^2)\ldots(1+P_l^2P^2)e_0g_1g_2\ldots g_l} \tag{36}$$

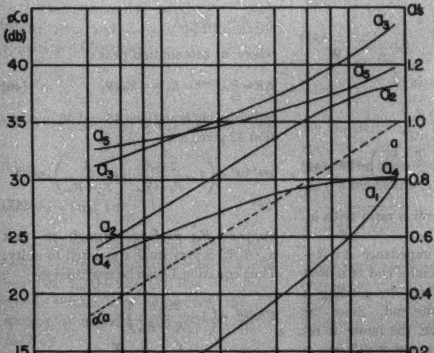


Figure 8. Element values of semifinite terminated filters designed on insertion loss basis. $k=0.866$, $P_1=0.664$, $P_2=0.908$

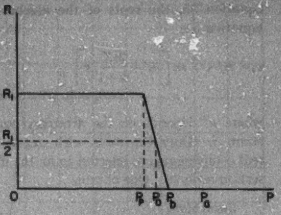


Figure 9. Approximate driving-point resistance of a ∞-1 filter

The expression given in equation 36 can be put in fraction form and compared with a similar expression as given in equations 12 and 18 to 22 inclusive for a semifinite terminated filter of n-elements ($n=2l+1$). By equating these two expressions, the coefficients $a_1, a_2 \ldots a_n$ can be determined.

The most serious difficulty in this procedure lies in the determination of the zeros of the power ratio. The equation one obtains by equating the right-hand side of equation 35 to zero is $2l+1$th order in P. To solve a high order algebraic equation with complex roots accurately is a very difficult task. One might be tempted to discard equation 35 altogether, and to use equation 36 instead as a starting point. However, such an apparent solution would not be satisfactory for the following reason: While almost any even choice of the constants P_1, $P_2, \ldots$ and $Q_1, Q_2, \ldots$ in equation 35 leads to a satisfactory filter, the constants e_0, h_1, $h_2 \ldots g_1, g_2, \ldots$ and P_1, $P_2 \ldots$ in equation 36 must be so well coordinated that a slight deviation of one constant from its proper value is likely to lead to excessive passband loss at certain frequencies. The roots of the algebraic equation, however, can be readily calculated for a special case. Darlington[2] has shown that for filters with equal maxima in the passband and equal minima in the attenuation band of the insertion loss curve, the constants in equation 35 are

$$P_i=Q_i=\sqrt{k}\,sn\left[\frac{2iK}{2l+1},k\right] i=1, 2 \ldots l \tag{37}$$

In equation 37, K is the value of the complete elliptic integral. The modulus k is the ratio of the upper frequency limit f_p of the passband to the lower frequency limit f_a of the attenuation band for a low-pass filter. For other types of filters, k is the corresponding ratio P_p/P_a of the frequency variable P, as defined in equation 1. With these constants in equation 35, the roots of the algebraic equation are

$$\pi_i = \pm j\sqrt{k}\, sn\left[ju' + \frac{2iK}{2l+1}, k\right], \qquad i = 0, \pm 1, \pm 2 \ldots \pm l$$

where u' depends on the arbitrary constant S. Only the roots with negative real components are selected to fix the insertion ratio 36. One obtains

$$h_i = \frac{2\epsilon_0 cnu_i dnu_i}{1+\epsilon_0^2 P_i^2} \tag{38}$$

$$g_i = \frac{\epsilon_0^2 + P_i^2}{1+\epsilon_0^2 P_i^2} \tag{39}$$

where

$u_i = 2iK/(2l+1)$
$i = 1, 2, 3 \ldots$
$\epsilon_0 = j\sqrt{k}\, sn\,(ju', k)$

The choice of u' determines the maximum value of insertion loss in the passband. The calculations are performed for the two section filters, and the results are plotted into curves as shown in Figures 5 to 8 inclusive. Each curve sheet corresponds to a value of k, or P_p/P_a.

From these curves, the capacitance and inductance values for various reactance arms, as well as the insertion loss in the attenuation band corresponding to a given maximum value of frequency distortion in the passband, can be readily determined. The element values are

$L_0 = \dfrac{R_1}{2\pi f_c}$ henrys

$C_0 = \dfrac{10^6}{2\pi f_c R_1}$ microfarads

$C_1 = a_1 C_0$

$C_2 = \dfrac{P_1^2}{a_2} C_0, \qquad L_2 = a_2 L_0$

$C_3 = a_3 C_0$

$C_4 = \dfrac{P_2^2}{a_4} C_0, \qquad L_4 = a_4 L_0$

$C_5 = a_5 C_0$

Minimum Ratio of Termination

In practice, the grid circuit of a vacuum tube as a filter termination is approximately an infinite impedance. The plate circuit of a pentode can be considered a constant current source. However, the input end of a shunt feedback amplifier in its simplest form is far from being an infinite conductance, and the cost goes up as the equivalent conductance is increased. Similarly, the output end of a shunt feedback amplifier approaches a constant voltage source only at the expense of increased cost. Therefore, it is necessary to determine the minimum ratio of terminating resistances for a given limit of distortion so introduced.

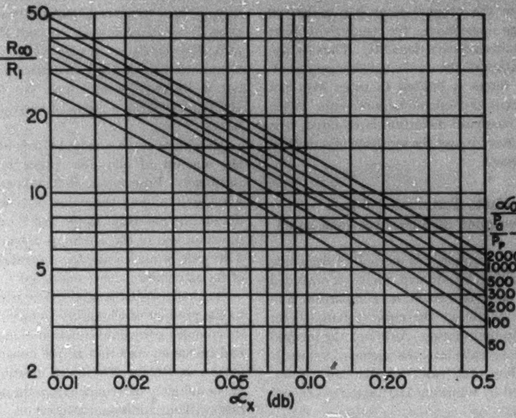


Figure 10. The minimum allowable ratio of terminating resistances

Suppose the output end of a 1-∞ filter is terminated in a resistance of R_∞. Looking from R_∞ into the filter, there is a voltage source E_2 in series with an impedance Z. E_2 is the open-circuit voltage and Z is the driving-point impedance at the output end.

The voltage E_2' developed across R_∞ is thus

$$E_2' = \frac{E_2 R_\infty}{Z+R_\infty} \tag{40}$$

Without the filter the voltage E_2'' would be

$$E_2'' = \frac{E_1 R_\infty}{R_1+R_\infty} \tag{41}$$

Hence the insertion ratio is

$$\epsilon^{\alpha'+j\beta'} = \frac{E_2''}{E_2'} = \frac{(Z+R_\infty)E_1}{(R_1+R_\infty)E_2} = \left(1+\frac{Z-R_1}{R_1+R_\infty}\right)\epsilon^{\alpha+j\beta} \tag{42}$$

where $\epsilon^{\alpha+j\beta}$ is the insertion ratio with infinite termination.

The driving-point impedance $Z = R + jX$ looking into the infinite end of a ∞-1 filter can be determined by sending a current into the infinite end. Since the filter is purely reactive, the power it receives is equal to the power it delivers to the terminating resistance R_1 at the finite end. Hence

$$|I_1^2|R = |I_2^2|R_1 \tag{43}$$

Since $|I_1^2/I_2^2| = \epsilon^{2\alpha}$ by definition, equation 43 can be written as

$$R = R_1\epsilon^{-2\alpha} \tag{44}$$

From equation 44, the resistive part of the driving-point impedance can be calculated from the power ratio. Since the terminating element of the infinite end is a shunt element, the driving-point impedance is of the minimum-reactance type. X can be calculated approximately from the expression for R. The method is given by H. W. Bode.[3]

In the passband, the points of maximum deviation of R from R_1 corresponds to the points of peak attenuation, and this deviation is very small. $Z - R_1$ can be written as

$$Z - R_1 = \Delta R + jX \tag{45}$$

where, at attenuation peaks

$$\Delta R = R_1\epsilon^{-2\alpha_p} - R_1 \doteq -2\alpha_p R_1 \tag{46}$$

Substituting equations 45 and 46 in equation 42 gives

$$\epsilon^{\alpha_p'+j\beta'} = \left(1 - \frac{2\alpha_p R_1}{R_1+R_\infty} + j\frac{X}{R_1+R_\infty}\right)\epsilon^{\alpha_p+j\beta} \tag{47}$$

Suppose R_∞ is large enough so that α_p/R_1R_∞ is very small compared to unity. Then equation 47 can be rewritten as

$$\epsilon^{\alpha_p'+j\beta'} \doteq \left(1-\frac{2\alpha_p R_1}{R_1+R_\infty}\right)\times \left(1+j\frac{X}{R_1+R_\infty}\right)\epsilon^{\alpha_p+j\beta} \tag{48}$$

The absolute magnitudes of the two sides of equation 48 are taken

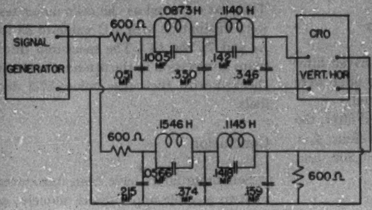


Figure 11 (left). Experimental verification of the equivalence between an image parameter filter and a semifinite terminated filter

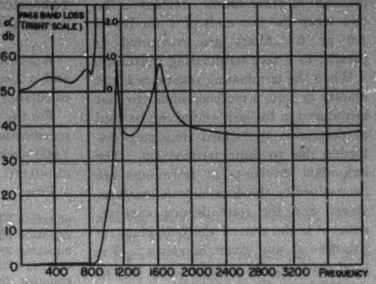


Figure 12 (right). Measured insertion loss of a 2-section semifinite terminated filter

$$\epsilon^{\alpha_p'}=\left(1-\frac{2\alpha_p R_1}{R_1+R_\infty}\right)\times \left(1+\left(\frac{X}{R_1+R_\infty}\right)^2\right)^{1/2}\epsilon^{\alpha_p} \quad (49)$$

Note that

$1-\frac{2\alpha_p R_1}{R_1+R_\infty}$ is approximately $\epsilon^{-\frac{2\alpha_p R_1}{R_1+R_\infty}}$

Let α_x be

$$\frac{1}{2}\log_e\left[1+\left(\frac{X}{R_1+R_\infty}\right)^2\right]$$

the equation 49 can be rewritten as

$$\alpha_p'=\frac{R_\infty-R_1}{R_\infty+R_1}\alpha_p+\alpha_x \quad (50)$$

If R_∞ is large enough so that $[X/(R_1+R_\infty)]^2$ is very small compared to unity

$$\alpha_x=\frac{1}{2}\left(\frac{X}{R_1+R_\infty}\right)^2 \text{ nepers} \quad (51)$$

It is therefore evident that the effect of a finite termination instead of an infinite termination is that while the individual attenuation peaks in the passband are somewhat reduced, an additional passband insertion loss is introduced by the reactance term.

In the attenuation band, R is practically zero, and $Z=jX$ approximately. Equation 42 becomes

$$\epsilon^{\alpha'+j\beta'}=\frac{R_\infty+jX}{R_1+R_\infty}\epsilon^{\alpha+j\beta} = \frac{R_\infty}{R_1+R_\infty}\left(1+\frac{jX}{R_\infty}\right)\epsilon^{\alpha+j\beta} \quad (52)$$

Thus

$$\alpha_a'\geq\alpha_a-\log_e\left(1+\frac{R_1}{R_\infty}\right) \quad (53)$$

Equation 53 shows that the insertion loss α_a' in the attenuation band is a little less than α_a, the insertion loss with open-circuit termination. However, the difference is never greater than $\log_e(1+R_1/R_\infty)$.

Since R_∞ is always large compared with R_1 as required to keep α_x small, the value of $\log_e(1+R_1/R_\infty)$ is usually less than 1 decibel, which is not very significant as compared to the insertion loss in the attenuation band.

To determine the value of X in the passband, the driving-point resistance R as a function of the frequency variable P is approximated by the solid line in Figure 9. P_0 is defined as the value of P at which the insertion loss is 3 decibels, and correspondingly $R=R_1/2$ at $P=P_0$. Let P_b be defined as the extrapolated value of P at which $R=0$. Then $P_b-P_p=2(P_0-P_p)$. A relationship between the real and imaginary components of a passive network function is given by H. W. Bode.[3] Though in the original derivation, the angular frequency ω is used as the independent variable, the same derivation holds true if the frequency variable P is used instead of ω. Likewise, we have

$$X_c=-\frac{R_1}{\pi(P_b-P_p)}\int_{P_p}^{P_b}\log_e\left|\frac{P+P_c}{P-P_c}\right|dP \quad (54)$$

where X_c is the value of X at $P=P_c$.

Equation 54 is integrated to give

$$X_c=-\frac{R_1}{\pi}\left\{\frac{P_b}{P_b-P_p}\log_e\left|\frac{P_b+P_c}{P_b-P_c}\right|-\frac{P_p}{P_b-P_p}\log_e\left|\frac{P_p+P_c}{P_p-P_c}\right|+\frac{P_c}{P_b-P_p}\log_e\left|\frac{P_b^2-P_c^2}{P_p^2-P_c^2}\right|\right\} \quad (55)$$

By differentiating equation 55 with respect to P_c and equating the resultant expression to zero, the value of P_c at which X_c is a maximum is

$$P_c=\sqrt{\frac{1}{2}(P_b^2+P_p^2)} \quad (56)$$

Apparently the maximum value of X occurs in the transition region. The highest value of X in the passband is at $P_c=P_p$. Equation 55 becomes

$$X_c=-\frac{R_1}{\pi}\left\{\frac{P_b+P_p}{P_b-P_p}\log_e\left(\frac{P_b+P_p}{2P_p}\right)-\log_e\left(\frac{P_b-P_p}{2P_p}\right)\right\} \quad (57)$$

Let v denote the ratio $(P_b-P_p)/2P_p$, then equation 57 can be written as

$$X_c=-\frac{R_1}{\pi}\left\{\frac{v+1}{v}\log_e(v+1)-\log_e v\right\} \quad (58)$$

In most cases, the values of P_b and P_a are not known. Assuming that the insertion loss α increases linearly between P_p and P_a, we have

$$P_0-P_p=\frac{3}{\alpha_a}(P_a-P_p)$$

and

$$v=\frac{3}{\alpha_a}\left(\frac{P_a}{P_p}-1\right) \quad (59)$$

The equations 51, 58, and 59 express the ratio of terminating resistances R_∞/R_1 as a function of the additional passband distortion α_p and the rate of rise of the attenuation curve $\alpha_a/[(P_a/P_p)-1]$. This ratio is calculated and plotted in Figure 10.

For 0-1 and 1-0 filters, the same curves of Figure 10 can be used with the ratio R_1/R_0 in the place of R_∞/R_1. R_0 is the terminating resistance at the zero end.

In actual applications, the terminating impedance at the infinite or zero end of the semifinite filters are usually not purely resistive. The shunt capacitance of a high-impedance termination, and the series reactance of a low-impedance termination, can be absorbed in the last elements of the 1-∞ and 1-0 filters respectively. The only component which needs to be considered is the resistive component.

Series and Parallel Connection of Channel Filters

Evidently, when the infinite ends of a number of 1-∞ or ∞-1 filters are connected in series to a terminating resistance of practically infinity, the performance of each filter is not influenced by the presence of other filters. Similar is the case when the zero ends of a number of 1-0 or 0-1 filters are connected in parallel to a zero terminating resistance.

When the terminating resistance is not infinity or zero, a method similar to that introduced in the last section can be used to determine the added frequency distortion due to finite termination. The individual driving-point resistances are determined by the discrimination requirements, and the total driving-point resistance is the sum of the individual resistances at any given frequency. The total driving-point reactance can be calculated from the total driving-point resistance. From these values, the added passband distortion can be calculated as in the preceding section.

For a group of filters with separate passbands, only one of the individual driving-point resistances can be of substantial magnitude for any given frequency. The others are negligible compared to the terminating resistance. The reactance components of adjacent filters are always opposite in sign, and the total reactance is lower than the reactance of each individual filter, except at the two ends of the total frequency band. If impedance compensation is used at the two ends, a terminating resistance would be high enough for all the filters in series if it is high enough for each individual without the presence of others. Thus the minimum value of the terminating resistance can be determined individually for each filter from the curves of Figure 10. The actual terminating resistance should be equal to or higher than each of the minimum allowable values as obtained from Figure 10.

For channel filters, it is usually desirable to ground every filter. This can be done for filters connected in parallel. For 1-∞ or ∞-1 filters connected in series, with not more than one low-pass filter in the group, the grounding of other filters can be done by replacing a coil with a transformer. The shunt reactance arm at the infinite end of a 1-∞ or a ∞-1 filter consists of a coil and a capacitor in parallel for a band-pass filter or a coil only for a high-pass filter. This coil can be replaced by a transformer with the same self-inductance in terms of the winding at the filter end. For ideal open-circuit termination the receiving-end voltage would be multiplied uniformly by the ratio of mutual inductance to self-inductance of the transformer. A transformer so introduced would not cause any additional frequency distortion.

Laboratory Studies of Open-Terminated Filters

A 2-section equivalent m-derived filter and a 2-section insertion loss filter were calculated, assembled, and tested.

For the equivalent m-derived filter, the specifications were: $R_1=600$ ohms, $f_c=1{,}000$ cycles per second, $m_1=0.809$, and $m_2=0.600$.

From

equation 23 $a_1=0.1914$
equation 26 $a_2=0.914$
equation 28 $a_3=1.320$
equation 30 $a_4=1.194$
equation 29 $a_5=1.306$

Then

$L_0=R_1/2\pi f_c=0.0955$ henry
$C_0=10^6/2\pi f_c R_1=0.2655$ microfarad
$C_1=0.051$ microfarad
$C_2=0.1005$ microfarad
$C_3=0.350$ microfarad
$C_4=0.142$ microfarad
$C_5=0.346$ microfarad
$L_2=0.0873$ henry
$L_4=0.1140$ henry

The receiving-end voltage of this filter was compared with that of an ordinary 2-section m-derived filter with the same specifications by applying them to the vertical and horizontal inputs of a cathode-ray oscilloscope as shown in Figure 11. The observed oscilloscope pattern was a straight line inclined at 45 degrees for an input frequency varying from 50 cycles per second up to 5,000 cycles per second. The result verified that the two receiving-end voltages were in phase and were proportional to each other as expected.

For the insertion loss filter, the specifications were: $R_1=600$ ohms, $f_c=1{,}000$ cycles per second, $f_p/f_a=0.800$, and $\alpha_p=0.30$ decibel. Then reading from the curves of element values

$a_1=0.57$
$a_2=1.17$
$a_3=1.34$
$a_4=1.015$
$a_5=1.23$

The actual elements were calculated to be

$C_1=0.151$ microfarad
$C_2=0.0845$ microfarad
$C_3=0.356$ microfarad
$C_4=0.196$ microfarad
$C_5=0.326$ microfarad
$L_2=0.1118$ henry
$L_4=0.0970$ henry

The measured results are shown in Figure 12. The measured attenuation in the passband was higher than the calculated value for pure reactive elements. This was expected as the coils under test were dissipative with a Q value of 35 for L_2 and 75 for L_4. The measured minimum attenuation in the attenuation band checked with the curve value, 37.3 decibels.

Conclusion

Semifinite filters have been introduced for two possible applications, namely, as filters inserted between a finite resistance and a very high or very low impedance, and as channel filters which are connected in series or in parallel at one end. For the former application up to 6-decibel gain is obtainable without additional cost, and for the latter application each filter can be designed individually without any complications due to the presence of other filters.

The minimum ratio of terminating resistances is found to be a function of the allowable passband frequency distortion and the sharpness of rise of the attenuation curve. This minimum ratio is well within practicable limits.

For an image parameter filter with prototype image impedance a corresponding semifinite filter can be assembled which, inserted between a constant resistance and an open or short circuit, gives the same attenuation and phase shift as does the original image parameter filter terminated with equal resistances.

The assumption of semifinite termination simplifies the process of computation when the insertion loss method of design is used. Curves for direct reading of element values have been prepared for the 2-section filters to meet moderate discrimination requirements.

References

1. CONSTANT RESISTANCE NETWORKS WITH APPLICATIONS TO FILTER GROUPS, E. L. Norton. *Bell System Technical Journal* (New York, N. Y.), volume 16, April 1937, pages 178–93.

2. SYNTHESIS OF REACTANCE 4-POLES, S. Darlington. *Journal of Mathematics and Physics* (Cambridge, Mass), volume 10, September 1939, pages 257–353.

3. NETWORK ANALYSIS AND FEEDBACK AMPLIFIER DESIGN (book), H. W. Bode. D. Van Nostrand Company, Inc., New York, N. Y., 1945.

4. THEORY AND DESIGN OF UNIFORM AND COMPOSITE ELECTRIC WAVE FILTERS, O. J. Zobel. *Bell System Technical Journal* (New York, N. Y.), volume 2, 1923, pages 1–46.

5. TRANSMISSION CHARACTERISTICS OF ELECTRIC WAVE FILTERS, O. J. Zobel. *Bell System Technical Journal* (New York, N. Y.), volume 3, 1924, pages 567–620.

No Discussion

The Effective Feedback Ratio of Magnetic Amplifiers

L. A. FINZI
MEMBER AIEE

G. F. PITTMAN, JR.
STUDENT MEMBER AIEE

H. L. DURAND
NONMEMBER AIEE

THE purpose of this paper is to present some approximate expressions relating the factors that govern the self-excitation in magnetic amplifiers. Possibilities are shown of a compensation of diversities of rectifier characteristics so as to obtain uniform steady-state performance, simply by readjustments of feedback turns, without changing any other factors of design or operation. Moreover, the effects of controlled or uncontrolled (drifting) variations of feedback circuit parameters are examined in terms of resulting current gains.

Two simplifying assumptions have been introduced in the treatment:

1. A rectifier is considered to be a constant resistor, R_d, for current in direct (forward) flow and as a constant (and usually much higher) resistor, R_r, for current in the reverse direction. Approximations of this kind seem almost unavoidable in any straightforward analysis aiming at the derivation of design formulas. In this problem values of R_d and R_r obtained from measurements of dynamic voltage-current characteristics of the rectifiers seem adequate to describe the rectifier behavior, at least for the power source frequency of 60 cycles used in the experimentation described later.
2. It is assumed that the control circuit has low impedance to the flow of even harmonic currents. The implications of this assumption are discussed in the Appendix.

The Magnetic Amplifier with External Feedback

This type of amplifier, shown in Figure 1(A), is considered first as it represents a most general type of circuit of which the doubler (without or with booster turns) and the full-wave circuit (also without or with booster turns) are particular cases in teams of circuit analysis, as shown in cross-reference tables given in reference 1.

Under the assumptions and approximations introduced in the Appendix, identity of instantaneous values of output current i_g results in the steady state for an original amplifier and for a second amplifier with modified feedback circuit parameters if the quantity

$$\xi=\left[\frac{N_f}{N_g}\,\frac{R_r-R_d}{R_r+R_d+4r_f+2R_f}\right] \qquad (1)$$

has the same numerical value in both instances. N_f and N_g are the turns of feedback and of gate circuits respectively linking one core; $2r_f$ is the total resistance of feedback windings; and R_f is a rectified resistive load introduced in the feedback loop for the sake of generality. It is specified, of course, that cores, gate, and control windings turns, N_g and N_c, and resistances, $2r_g$ and $2r_c$, d-c signal voltage V_c, power gate voltage v_g, and gate circuit operational impedance Z_g are kept unchanged in the two cases.

Although N_g is not a variable of the problem, it has been included in the expression of ξ which can be written then in the form

$$\xi=\lambda\sigma$$

where $\lambda=N_f/N_g$ is the "nominal" feedback ratio, and

$$\sigma=\frac{R_r-R_d}{R_r+R_d+4r_f+2R_f}$$

is a factor indicative of the reduction of feedback effectiveness which results from the leakage of any less than perfect rectifiers, and expresses the ratio of half-cyclic averages $\sigma=I_f/I_g$.

Various attempts have been made in the literature[2,3] to relate gains and general performance to the ratio λ; however, ξ is certainly a more significant and comprehensive quantity for any such purposes, and the name "effective feedback ratio" is suggested for it.

Variations of σ (for example, because of the use of different batches of rectifiers) may be compensated for by readjustments of λ. Also reductions of gain, if desired, can be obtained, in a given amplifier, either by reducing N_f or by reducing R_r by means of resistors in parallel with each rectifier, with identical steady-state performance in either case.

Also it may be noted that an amplifier with $\sigma<1$ can be thought of in terms of an amplifier with supposedly perfect rectifier cells, but with the rectifier bridge shunted by a single resistor R connected across the points c and d where

$$R=\frac{\sigma}{1-\sigma}(2r_f+R_f)$$

While it is established that equality of ξ's seems sufficient to insure identity of performance under the assumptions and approximations used, a second problem of practical importance arises at this point. Namely, one may want to know to what extent controlled or uncontrolled (drifting) variations of ξ will affect the amplifier characteristics.

Define the ampere-turn gain (or current gain on a common turn basis) as the slope $S=d(N_gI_g)/d(N_cI_c)$ of the transfer characteristic at the output under consideration, where I_g and I_c represent half-cyclic averages of output and control current. It is shown in the Appendix that the gains S' and S'' of two amplifiers, identical in all other respects of design and operating specifications but having different effective feedback ratios ξ' and ξ'', are related by

---

Paper 52-124, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE North Eastern District Meeting, Binghamton, N. Y., April 30–May 2, 1952. Manuscript submitted February 4, 1952; made available for printing April 15, 1952.

L. A. Finzi and G. F. Pittman, Jr., are with Carnegie Institute of Technology, Pittsburgh, Pa., with which H. L. Durand was associated at an early stage of this work.

This work was sponsored by the United States Office of Naval Research. Useful comments by W. F. Horton of Westinghouse Electric Corporation and by A. A. Sterk of Magnetic Amplifiers, Inc., are gratefully acknowledged.

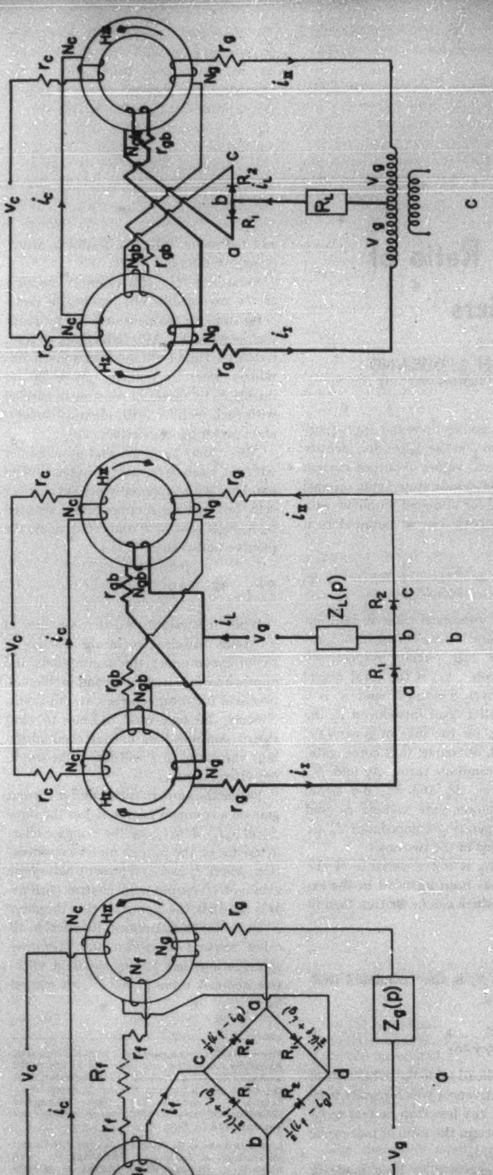


Figure 1(A). External feedback circuit. (B). Doubler circuit. (C). Full-wave circuit

$$\frac{S''}{S'}=\frac{1}{1+(\xi'-\xi'')S'} \text{ at any given } I_g \qquad (2)$$

The question of sensitivity to drifts of any of the factors affecting ξ can be evaluated by examining the quantity $D=(S'-S'')/S'$ where S' and S'' here are the gains before and after drift. For instance, it is seen immediately that lower values of D result from the use of temperature stable resistors shunting the rectifying elements, as stated by F. S. Malick in an unpublished paper presented at the AIEE Middle Eastern District Meeting in October, 1950. In fact, this procedure lowers S to start with and also makes σ (and thus ξ) less sensitive to drifts of R_r proper. Mr. Malick stated that a given percentage change of R_r has a smaller effect on gain when R_r is high. This may be somewhat less obvious, in view of the fact that amplifiers with higher R_r have a higher S to start with and therefore appear more sensitive to drifts of ξ. A comparison of the expressions of D for two amplifiers (α) and (β) with different rectifiers, but otherwise identical, proves, however, that for changes ΔR_r satisfying the condition $(\Delta R_r/R_r)_\alpha=(\Delta R_r/R_r)_\beta$ this statement is correct in all practical instances.

Some fallacies that may result from hasty judgments also can be dispelled by a careful consideration of the implications of the formulas. For example, one may suspect that a stabilization without the loss of gain which is inherent to the procedure just mentioned could be achieved easily by first paralleling the rectifiers with temperature stable resistors R_s and then increasing N_f so as to obtain a value $\xi^*=\lambda^*\sigma^*$ equal to the original $\xi=\lambda\sigma$ (and therefore also the original gain). Actually it is seen that the effect of a temperature drift of R_r upon $R_r^*=R_rR_s/(R_r+R_s)$ is certainly much less pronounced percentage-wise, but the ratio $d\xi^*/dR_r$: $d\xi/dR_r$ resulting may be larger than unity. In such cases the gain of the supposedly "stabilized" amplifier actually may be more sensitive to drifts of R_r proper.

(Stabilization against drift, of course, may be accomplished by shunting the rectifiers with suitably temperature sensitive resistors or networks so as to obtain R_r^* truly constant. By doing so σ^* may result to have an undesirably low value. Here, however, readjustments of gain can be made further by readjustments of λ with a net stabilization result.)

Self-Feedback Amplifiers with A-C Output

This type of amplifier, the so-called "doubler" type, is presented in Figure 1(B) in a form somewhat generalized by the addition of boosting turns N_{gb}. (To avoid duplication of writing it shall be established that N_{gb} is a positive number if these turns are connected with the polarities shown in the drawing so as to give magnetomotive force of direct conduction concordant with the magnetomotive force of direct conduction of the N_g windings on the same core. If in the use of this circuit the polarities of the N_{gb} windings are reversed, for example, to reduce and stabilize gains, it shall be understood that N_{gb} becomes inherently a negative number. In the conventional doubler, of course, $N_{gb}=0$.)

With the same assumptions of low control circuit impedance and semilinearized rectifier characteristics used before, the analysis developed in the Appendix shows that identity of instantaneous values of output i_L for two amplifiers with unchanged control circuit and voltage and unchanged gate load $Z_L(p)$, power source voltage, and frequency, but with modified other parameters of design, is practically obtained if the two expressions

$$\xi=\frac{(N_g+N_{gb})}{(N_g-N_{gb})}\,\frac{R_r-R_d}{R_r+R_d+2r_g+2r_{gb}} \qquad (3)$$

and

$$N_G=N_g-N_{gb} \qquad (4)$$

have the same numerical values in the two amplifiers under comparison. r_g and r_{gb} are the resistances of the gate and booster windings per core. The additional condition (equation 4) expresses a fact recognized first by Krabbe[4] as he states that "the active number of turns" or the turns that "act as the main winding" are actually N_g-N_{gb} with the symbols used here.

Again one can say that the effective feedback ratio defining the amplifier performance is

$$\xi=\lambda\sigma$$

where $\lambda=N_F/N_G$ (with $N_F=N_g+N_{gb}$ and $N_G=N_g-N_{gb}$), and

$$\sigma=\frac{R_r-R_d}{R_r+R_d+2r_g+2r_{gb}}$$

is a factor which is unity for "perfect" rectifiers and in any other case is indicative of the reduction of feedback action that results from rectifier leakage. Also $\sigma=(I_\mathrm{I}+I_\mathrm{II})\div(I_\mathrm{I}-I_\mathrm{II})$ where I_I and I_II are half-cyclic averages of the branch currents with the sign conventions shown in Figure 1(B).

It may be noted that an amplifier of this type with $\sigma<1$ can be thought of in terms of an amplifier with supposedly perfect rectifiers, but having a shunting resistor

$$R=\frac{\sigma}{1-\sigma}(2r_g+2r_{gb})$$

connected across the points a and c.

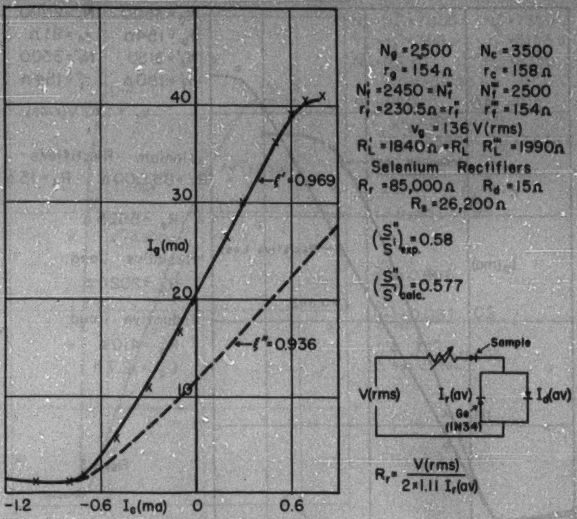


Figure 2

While identities of ξ and of N_G are sufficient to insure identity of performance (and in particular of gains at any given output), the further problem of the influence of controlled or uncontrolled (drifting) variations of ξ upon the slopes of the transfer characteristic arises at this point. Defining the ampere-turn gain as the slope of the transfer characteristic $S=d(N_GI_L)/d(N_cI_c)$ at a given output I_L (half-cyclic average) with the same procedures used in the preceding section it is found that the gains S' and S'' of two amplifiers identical in all other respects but having different effective feedback ratios ξ' and ξ'' are related by

$$\frac{S''}{S'}=\frac{2}{2+(\xi'-\xi'')S'} \text{ at any given } I_L \qquad (5)$$

Self-Feedback Amplifiers with Self-Rectified Output

This type of amplifier, the so-called "full-wave" type, is presented in Figure 1(C) also in a form somewhat generalized by the addition of turns N_{gb}. (As in the preceding section, N_{gb} is a positive number for boosting connection of these turns as shown. If the polarities of the N_{gb} windings are reversed it shall be understood that N_{gb} is inherently a negative number. In the conventional full-wave circuit, of course, $N_{gb}=0$.)

With the same assumptions and approximations used in the previous sections it is found here that identity of performance for two amplifiers with unchanged control circuit and voltage and unchanged resistive load R_L, power source voltage, and frequency but with modified other parameters of design is practically obtained if the two quantities

$$\xi=\frac{N_g+N_{gb}}{N_g-N_{gb}}\,\frac{R_r-R_d}{R_r+R_d+2r_g+2r_{gb}+4R_L} \qquad (6)$$

and

$$N_G=N_g-N_{gb} \qquad (7)$$

have the same numerical value in the two amplifiers under comparison. It should be pointed out that identity of performance here means identity of instantaneous values of the difference of branch currents $i_\mathrm{I}-i_\mathrm{II}$, which is slightly different from the identities of output currents obtained in the previous cases. Also, in this type of circuit, somewhat more serious limitations may arise for the practical validity of the criteria (equations 6 and 7) as seen in the Appendix in connection with equation 43. Again one can say that the effective feedback ratio is

$$\xi=\lambda\sigma$$

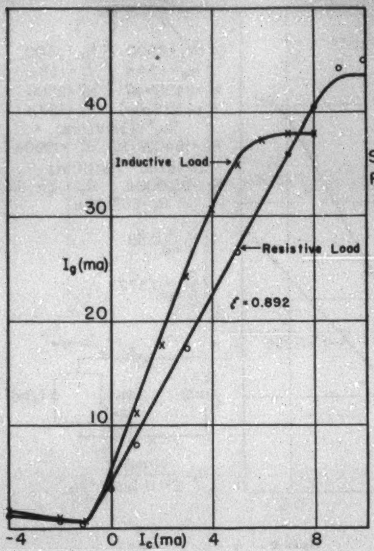


Figure 3

where $\lambda = N_P/N_G$ (with $N_P = N_g + N_{gb}$ and $N_G = N_g - N_{gb}$), and

$$\sigma = \frac{R_r - R_d}{R_r + R_d + 2r_g + 2r_{gb} + 4R_L}$$

is the factor embodying the influence of rectifier leakage.

Again

$$\sigma = \frac{I_I + I_{II}}{I_I - I_{II}}$$

and it may be noted also that an amplifier of this type with $\sigma < 1$ can be thought of in terms of an amplifier with supposedly perfect rectifiers, but having a shunting resistor

$$R = \frac{\sigma}{1-\sigma}(2r_g + 2r_{gb} + 4R_L + 4R_d)$$

connected across the points *a* and *c*.

The problem of the influence of controlled or uncontrolled variations of ξ upon the slopes of the transfer characteristics is treated as in the previous cases. Defining the ampere-turn gain as slope

$$S = \frac{d(N_G I_L)}{d(N_c I_c)}$$

of the transfer characteristic at output I_L (average), it is shown in the Appendix that the gains S' and S'' of two otherwise identical amplifiers with different effective feedback ratios are related by

$$\frac{S''}{S'} = \frac{2\sigma''}{2\sigma' + (\xi' - \xi'')S'} \tag{8}$$

where the slopes S' and S'', in this case, are compared at two slightly different ordinates of the two transfer characteristics, namely at I_L' and at $I_L'' = I_L'\sigma''/\sigma'$.

Experimentation

A broad variety of tests were performed at the power source frequency of 60 cycles on amplifiers with Deltamax and Hypernik V cores and also on an amplifier with a less sensitive core material (Hypersil). Also different types of rectifiers were used. For these the determination of R_r was made from dynamic tests of the type indicated in Figure 2, on the basis of a voltage V_{rms} having the same crest value as the peak of reverse voltage expected to appear across the rectifier at full output. The influence of varying I_d was found to be small over a broad range. R_d was obtained similarly from the voltage drop caused by sinusoidal half-wave currents.

Figures 2 to 5 refer to tests performed on external feedback circuits with Deltamax cores. In Figure 2 the characteristic

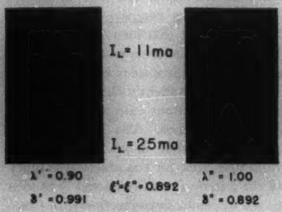
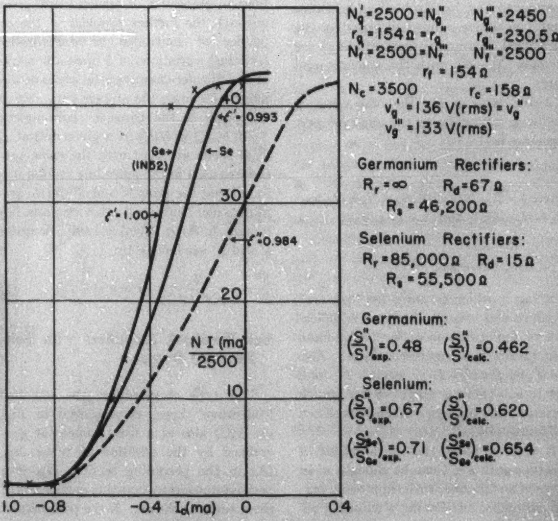


Figure 4. Wave forms of output currents

Figure 5 (right)



of the original amplifier with $\lambda' = 0.98$ in full line was modified into the dotted characteristic shown by increasing the leakage by means of resistors R_s shunting each arm of the rectifier bridge. The reduction of σ thus achieved then was compensated by increasing N_f, and thus λ, so as to re-establish the original ξ. The crossed points thus obtained fall on the original characteristic as expected. Also the calculated ratio of gains around the center of design for the full line and dotted line characteristics checked well with the experiment.

Figure 3 shows similar tests performed on the original amplifier with $\lambda' = 0.90$ with a purely resistive load and also with a highly inductive load. Again the bridge arms were shunted by resistors and then N_f was increased so as to re-establish the original ξ. Oscillograms of output current wave forms were also taken to check on the expected identity of instantaneous values of output current for amplifiers with identical ξ's. Such oscillograms are shown in Figure 4 for the tests of Figure 3 with inductive load.

Figure 5 shows in full line characteristics obtained for an amplifier with $\lambda' = 1.00$, with selenium and with germanium rectifiers. The calculated loss of gain using the selenium rectifiers is checked experimentally. The dotted line shows the characteristic obtained with the same λ, and germanium rectifiers, but with resistors R_s shunting each arm of the bridge. The measured loss of gain was slightly less than predicted. A subsequent increase of N_f so as to re-establish the original ξ resulted in a slight overcompensation, as shown by the crossed points. The amplifier with selenium rectifiers also was submitted to a similar artificial increase of leakage and subsequent compensation by increase of N_f, but a more pronounced overcompensation resulted. Exact calculations of N_f can hardly be expected in this very critical region, but any overcompensation can be corrected in a further step by a suitable shunting resistor across the d-c terminals of the bridge. Similar tests with $\lambda' = 0.80$ and $\lambda'' = 1.20$ were performed with good agreement.

Figure 6 shows tests on the same amplifier in doubler connection. The dotted line is the characteristic of the amplifier with $\lambda' = 1.00$. The full line is the characteristic of the same amplifier with $\lambda'' = 0.90$. The crossed points apply to the amplifier with $\lambda''' = 1.00$ but with σ reduced (by shunting each rectifying element with a resistor) so as to obtain $\xi''' = \xi''$. Checks of gains were also satisfactory.

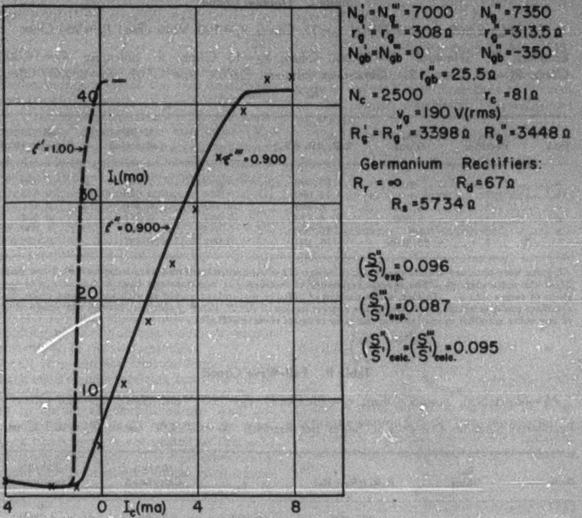


Figure 6

Further tests were performed on the Hypernik V amplifier with $\lambda = 1.00$ throughout, and calculated ratios of gains for various R_r (resulting from shunting resistors or also from the use of rectifiers with different characteristics) were checked against the experimental results, as shown in Tables I and II.

A variety of similar tests were performed on the Hypersil amplifier. The agreements here were equally good or better. The lesser over-all sensitivity of this device seemingly overcomes the fact that some of the approximations used in the analysis may appear somewhat more dubious in the case of these core materials.

In the tests of Tables I and II rms as well as rectified average outputs were recorded and the same agreement between rms characteristics was found. This confirms the identity of instantaneous values evidenced by the oscillograms of Figure 4.

Finally tests were performed to gather information on the limitations of the assumption of low control circuit resistance. For this forcing resistors were connected in series with the control windings so as to increase the control circuit resistance by factors of 8, or even 15 in some cases. In these cases, the transfer characteristics differed somewhat from those obtained without forcing resistors, but the same agreement was found between characteristics for different σ but equal ξ and also between calculated and experimental gain ratios. On the other hand, no such agreement was found in other tests in which the basic assumptions were grossly violated by the use of a choke of about 1,800 henrys in series in the control circuit.

Conclusions

Formulas have been presented and checked with careful experimentation which make evident the separate influence of parameters of design affecting the self-excitation of magnetic amplifiers with low control circuit impedance. Possibilities of compensation of unwanted effects of some factors by means of readjustments of others are shown, and the influence of controlled or uncontrolled variations of the various parameters upon current gains is expressed.

Appendix

External Feedback Circuit: Figure 1(A)

Kirchhoff's voltage equations for control, gate, and feedback loops are

$$v_c = AN_c\left(\frac{dB_I}{dt} + \frac{dB_{II}}{dt}\right) + 2r_c i_c \qquad (9)$$

Table I. Doubler Circuit

$N_g = N_c = 2{,}500$; $r_g = 65$ Ohms, $r_c = 75$ Ohms; $V_g = 120$ Volts (Rms) $R_L = 964$ Ohms

Rectifiers: 1. Selenium: $R_r = 85{,}000$ Ohms, $R_d = 15$ Ohms; 2. Selenium: $R_r = 14{,}000$ Ohms, $R_d = 15$ Ohms; 3. Germanium (General Electric type G10): $R_r = 85{,}000$ Ohms, $R_d = 5$ Ohms

Run	Rectifier	RR_s, Ohms	$R_rR_s/(R_r+R_s)$	σ	(S″/S′) Calculated	(S″/S′) Experimental
a	1	∞	85,000	0.998	1.000	1.000
b	2	∞	14,000	0.990	0.638	0.643
c	1	10,000	8,950	0.982	0.466	0.451
d	1	5,000	4,720	0.967	0.311	0.304
e	3	∞	85,000	0.998	1.000	1.000
f	3	20,300	16,400	0.992	0.667	0.630

All gains were obtained around the center of design ($I_L = {}^1/_2 I_L$ saturated) and were compared with the gain S' of run *a* throughout. The reverse resistance of rectifiers (3) was obtained from a comparison of the gains of runs *e* and *a*. (In fact, a comparison of gains of a standard amplifier with known and unknown rectifiers could be taken as a basis for the determination of the R_r value of rectifiers used in the production of any other amplifier operating over the same range of reverse voltages.)

Table II. Full-Wave Circuit

$N_g = N_c = 2{,}500$; $r_g = 65$ Ohms, $r_c = 75$ Ohms; $V_g = 120$ Volts (Rms); $R_L = 988$ Ohms

Rectifiers: Selenium (12 cells in Series per Branch); $R_r = 375{,}000$ Ohms, $R_d = 105$ Ohms

Run	R_s, Ohms	$R_rR_s/(R_r+R_s)$	σ	(S″/S′) Calculated	(S″/S′) Experimental
a	∞	375,000	0.989	1.000	1.000
b	680,000	242,000	0.983	0.831	0.815
c	340,000	178,500	9.976	0.693	0.705
d	250,000	150,000	0.972	0.632	0.675
e	194,000	128,000	0.967	0.570	0.574

S′ in all cases is gain of run *a*.

$$v_g = AN_g\left(\frac{dB_{\rm I}}{dt} - \frac{dB_{\rm II}}{dt}\right) + v_{ad} - v_{bd} + 2r_gi_g + Z_g(p)i_g \quad (10)$$

$$0 = AN_f\left(\frac{dB_{\rm I}}{dt} + \frac{dB_{\rm II}}{dt}\right) + v_{cb} + v_{bd} + 2r_fi_f + R_fi_f \quad (11)$$

with self-explanatory notation. B is the instantaneous value of flux density and A is the cross section of either one of the cores I or II. Assuming identical rectifiers in all bridge arms the voltage drops around the bridge are

$$v_{ad} = v_{cb} = R_1\frac{i_f + i_g}{2} \text{ and } v_{ca} = v_{bd} = R_2\frac{i_f - i_g}{2}$$

where R_1 and R_2 are the rectifier resistances of adjacent sides of the bridge at the instant considered. It is understood that such resistances have the value R_d (or respectively R_r) in any one rectifier if the voltage drop is positive (or respectively negative).

Magnetomotive force equations are

$$lH_{\rm I} = N_ci_c + N_fi_f + N_gi_g \quad \text{for core I} \quad (12)$$

$$lH_{\rm II} = N_ci_c + N_fi_f - N_gi_g \quad \text{for core II} \quad (13)$$

where l is the mean length of the magnetic path and H's are field intensities.

Magnetic characteristics of cores (assumed matched) are

$$B_{\rm I} = f(H_{\rm I}) \text{ and } B_{\rm II} = f(H_{\rm II}) \quad (14)$$

Where f is not necessarily a single-valued function but represents actually the existing relation between B and H as determined by the core history.

Two magnetic amplifier systems with comparable core materials but of otherwise different design yield proportionality of instantaneous values of currents in corresponding circuits if certain conditions of similitude are met which relate data of design and voltages applied in the two cases, as shown in reference 1. The problems approached in this paper, however, are characterized by a number of restrictions as it is desired primarily to compensate for diversities of feedback rectifier leakages so as to obtain identity of outputs either by means of readjusted feedback turns or by means of readjusted average control currents only while all other data of design and operation of the systems are kept unchanged. This leads to a number of contradictions in the conditions expressed by the general rules of similitude, that is—mathematically speaking—no such simple compensation is obtainable under the restrictions stated.

From an engineering viewpoint, however, it is legitimate to inquire whether some of the contradictory requirements of the rigorous model theory may be less significant than others and may be disregarded practically. Some indications may be obtained from the available amount of factual information concerning orders of magnitude and actual significance of various effects recognizable in the operation of amplifiers of common design. Ultimately, of course, the results of such speculations must be proved (or disproved) in their validity by careful experimentation. This search is performed in the following for the steady state, with an impressed d-c signal voltage V_c under the assumption commonly stated by saying that the control circuit has "low impedance to the flow of even harmonic currents."

(A) Since the problems arise from the fact that different rectifiers affect the values of feedback currents corresponding to a given output, a first step shall be to derive some simple expression relating feedback currents to rectifier "resistances" and other pertinent factors. Equation 9 is rewritten

$$0 = AN_c\left(\frac{dB_{\rm I}}{dt} + \frac{dB_{\rm II}}{dt}\right) + 2r_ci_{ch} \quad (15)$$

where i_{ch} is the harmonic content of control current $i_{ch} = i_c - I_c$ and $I_c = V_c/2r_c$ is the average control current. Substituting in equation 11

$$0 = -\frac{N_f}{N_c}2r_ci_{ch} + v_{cb} + v_{bd} + 2r_fi_f + R_fi_f \quad (16)$$

It is seen that if r_c were truly zero, v_{cb} and v_{bd} certainly could not be positive simultaneously, that is, only two (nonadjacent) rectifiers of the bridge could be in forward conduction at any time in the cycle. Commutation would not occur and the statement would be strictly true that

$$\text{either } R_1 = R_d \text{ and } R_2 = R_r \text{ or } R_1 = R_r \text{ and } R_2 = R_d \quad (17)$$

Also the reverse voltage across a nonconducting rectifier is then

$$v_r = -(2r_f + R_f)i_f - R_d\frac{|i_g| + i_f}{2} \cong (2r_f + R_f + R_d)|i_g|$$

It remains to be seen whether equation 17 can be accepted in the way of reasonable approximation in actual cases in which the control circuit resistance may be low, but certainly not quite zero. For the sake of simplicity let us consider the instance of an amplifier in which copper volumes, mean turn lengths, and number of turns are equal for control and feedback windings. Then in the absence of forcing resistors in the control circuit $r_c = r_f$ and equation 16 becomes

$$0 = v_{cb} + v_{bd} + 2r_f(i_f - i_{ch}) + R_fi_f$$

Now in amplifiers with commonly sensitive core materials and a reasonably large amount of positive feedback i_{ch} is algebraically quite small compared to i_f. Even if $R_f = 0$, the eventuality of v_{cb} and v_{bd} being simultaneously positive for any significant length of time during the cycle may be disregarded. (In fact, it may still be disregarded even if $2r_c$ is increased, for example, by the use of some moderate forcing resistor.) Making use of equation 17 in equation 11 and integrating over one half-cycle from an instant, t_0, of transition of i_g through zero

$$0 = \frac{2}{T}\int_{t_0}^{t_0+T/2}\frac{(dB_{\rm I} + dB_{\rm II})}{dt}dt + R_d\frac{I_f + I_g}{2} + R_r\frac{I_f - I_g}{2} + (2r_f + R_f)I_f \quad (18)$$

where T is the period and I_f and I_g are half cyclic averages (rectified) of feedback and gate (output) currents. Because of symmetry $B_I(t)=B_{II}(t+T/2)$; hence

$$\int_{t_0}^{t_0+T/2}\frac{(dB_I+dB_{II})}{dt}\,dt=0$$

and equation 16 yields

$$\frac{I_f}{I_g}=\frac{R_r-R_d}{R_r+R_d+4r_f+2R_f}=\sigma \tag{19}$$

The quantity σ is indicative of the reduction of average feedback current brought about by any less than infinite rectifier reverse resistance.

(B) A second step is to show that two amplifiers with identical cores and gate windings and identical impressed gate voltage v_g and gate load impedance Z_g, but with different feedback circuit parameters (feedback windings, feedback loop resistances, and rectifiers), yield identical instantaneous values of output current, i_g, if the quantity $N_cI_c+N_fI_f$ has the same numerical value in both instances. (I_c and I_f are cyclic averages of control and feedback current.)

For this equation 10 is rewritten eliminating i_f by means of equation 11 using equation 17

$$v_g=AN_g\left(\frac{dB_I}{dt}-\frac{dB_{II}}{dt}\right)+\rho i_g+AN_f\left(\frac{dB_I}{dt}+\frac{dB_{II}}{dt}\right) \tag{20}$$

where

$$\rho=\left[(2r_g+Z_g)+\left\{\frac{R_r(2r_f+R_f+2R_d)+R_d(2r_f+R_f)}{R_r+R_d+4r_f+2R_f}\right\}\right] \tag{21}$$

and ρi_g is evidently the voltage "not absorbed" by the cores.

Under the assumption of low control circuit impedance the last term on the right side of equation 20 is considered small in comparison to the one or the other of the first two.

Moreover, writing

$$H_I=H_{0I}+H_1(t) \text{ and } H_{II}=H_{0II}+H_2(t) \tag{22}$$

(where by integration of equations 12 and 13 over a cycle

$$lH_{0I}=lH_{0II}=lH_0=N_cI_c+N_fI_f)$$

and using equation 14 with the notation

$$f'(x)=\frac{d}{dx}f(x)$$

equation 20 is modified further into

$$v_g=AN_g\left[f'(H_0+H_1(t))\frac{dH_1}{dt}-f'(H_0+H_2(t))\frac{dH_2}{dt}\right]+\frac{\rho l}{2N_g}(H_1-H_2) \tag{23}$$

A further relation between H_I and H_{II} results from the assumption of low control circuit impedance, namely

$$\frac{dB_I}{dt}=-\frac{dB_{II}}{dt}$$

that is

$$f'(H_0+H_1(t))\frac{dH_1}{dt}=-f'(H_0+H_2(t))\frac{dH_2}{dt} \tag{24}$$

Identity of H's at corresponding times in corresponding cores of the two amplifiers, and therefore identity of $i_g=(l/2N_g)(H_I-H_{II})$, results approximately if H_0 and ρ are numerically equal in the two cases (V_g, AN_g, l, and the core characteristics being identical in the problem at hand). Actually equation 19 shows that diversity of feedback circuit parameters may result in some small differences in the two values of ρ. The prescribed invariance of ρ can be achieved by some further readjustments of the resistive component of load impedance Z_g, but such minor readjustments evidently are of negligible effects upon the output current. This permits the statement that identity of $lH_0=N_cI_c+N_fI_f$ alone is sufficient to insure identity of instantaneous values of output current, i_g, in the two amplifiers.

(C) It is seen now that reduction of feedback effectiveness resulting from the use of poorer rectifiers can be compensated for so as to obtain unchanged output with unchanged control, N_cI_c. For this the feedback turns must be readjusted so that σN_f or, if one prefers, $(N_f/N_g)\sigma=\xi$ is numerically equal in the amplifiers under comparison. This is the condition which is given in equation 1.

(D) If, on the other hand, no such readjustment of feedback design data is contemplated, the identity of lH_0 yielding identity of output with different σ's can be achieved only if N_cI_c is changed.

Let

$$S'=\frac{d(N_gI_g)}{d(N_cI_c')}$$

be the ampere-turn gain of a given amplifier (with $\xi=\xi'$) at a certain rectified average output I_g, and let

$$S''=\frac{d(N_gI_g)}{d(N_cI_c'')}$$

be the gain of another amplifier, identical to the first one but for the feedback circuit (so that $\xi=\xi''$), at the same output I_g.

Rewriting

$$lH_0=N_cI_c'+\xi'N_gI_g=N_cI_c''+\xi''N_gI_g$$

forming the ratio of the two gains

$$\frac{S'}{S''}=\frac{d(N_gI_g)/d(N_cI_c')}{d(N_gI_g)/d(N_cI_c'')}$$

and eliminating N_cI_c'' yields

$$\frac{S'}{S''}=1+(\xi'-\xi'')S'$$

at any given I_g, from which equation 2 is obtained.

Doubler Circuit: Figure 1(B)

(A) Kirchhoff's voltage equations are

$$v_c=AN_c\left(\frac{dB_I}{dt}+\frac{dB_{II}}{dt}\right)+2r_ci_c \tag{9}$$

$$v_g=A\left(N_g\frac{dB_I}{dt}+N_{gb}\frac{dB_{II}}{dt}\right)+(r_g+r_{gb})i_I+v_{ab}+Z_L(p)i_L \tag{25}$$

$$v_g=-A\left(N_{gb}\frac{dB_I}{dt}+N_g\frac{dB_{II}}{dt}\right)-(r_g+r_{gb})i_{II}-v_{bc}+Z_L(p)i_L \tag{26}$$

where I_I and i_{II} are the gate branch currents shown. With R_1 and R_2 rectifier resistances at the instant considered: $v_{ab}=R_1i_I$ and $v_{bc}=R_2i_{II}$.

The following notation is introduced

$$N_F=N_g+N_{gb},\ N_G=N_g-N_{gb}, \text{ and}$$

$$i_F=\frac{i_I+i_{II}}{2},\ i_G=\frac{i_I-i_{II}}{2}={}^1/_2iL \tag{27}$$

Thus the magnetomotive force equations are

$$lH_I=N_cI_c+N_Fi_F+N_Gi_G \quad \text{for core I} \tag{28}$$

$$lH_{II}=N_ci_c+N_Fi_F-N_Gi_G \quad \text{for core II} \tag{29}$$

From equations 25 and 26 by subtraction

$$0=AN_F\left(\frac{dB_I}{dt}+\frac{dB_{II}}{dt}\right)+v_{ab}+v_{bc}+2(r_g+r_{gb})i_F \tag{30}$$

which is quite similar to equation 11. Rewriting equation 9 in the form of equation 15 for control harmonic content with constant control voltage and substituting in equation 30

$$0=-\frac{N_F}{N_c}2r_ci_{ch}+v_{ab}+v_{bc}+2(r_g+r_{gb})i_F \tag{31}$$

which, in the ideal low control impedance case ($r_c=0$), excludes the possibility of simultaneous conduction of the two rectifiers as v_{ab} and v_{bc} could not be both positive simultaneously.

In this case the reverse voltage across one rectifier is found to be

$$v_r=-2(r_g+r_{gb})i_F-R_d(i_F+|i_G|)\cong-(r_g+r_{gb}+R_d)|i_L|$$

Even in more realistic situations, for reasonable winding designs and for commonly occurring low values of the ratio $N_F/N_ci_{ch}:2i_F$ the low control circuit impedance implies that the eventuality of v_{ab} and v_{bc} being simultaneously positive for any significant length of time during the cycle can be disregarded and equation 17 can be accepted tentatively. Integrating equation 30 over one half-cycle from an instant of transition of i_L through zero yields then, for half-cycle rectified average values

$$\frac{I_F}{I_G}=\frac{R_r-R_d}{R_r+R_d+2r_g+2r_{gb}}=\sigma \tag{32}$$

which is indicative of the reduction of average feedback action brought about by any less than infinite rectifier reverse resistance.

(B) Moreover, by addition of equations 25 and 26

$$v_g={}^1/_2AN_G\left(\frac{dB_I}{dt}-\frac{dB_{II}}{dt}\right)+\rho i_G+{}^1/_2AN_F\left(\frac{dB_I}{dt}+\frac{dB_{II}}{dt}\right) \tag{33}$$

where

$$\rho=\left[(r_g+r_{gb}+2Z_L)+\left\{R_r\frac{(r_g+r_{gb}+2R_d)+R_d(r_g+r_{gb})}{R_r+R_d+2r_g+2r_{gb}}\right\}\right] \tag{34}$$

Use of the assumption of low control circuit impedance is made to drop the last term on the right side of equation 33, which is rewritten using equation 22

$$v_g = {}^1/_2 A N_G \left[f'[H_0 + H_1(t)] \frac{dH_1}{dt} - f'[H_0 + H_2(t)] \frac{dH_2}{dt} \right] + \frac{\rho l}{N_G}(H_1 - H_2) \quad (35)$$

which, in conjunction with equation 24 indicates that identity of H's at corresponding times in corresponding cores (and therefore identity of $i_L = 2i_G = (l/N_G)(H_1 - H_2)$ results approximately if H_0, $N_G = N_g - N_{gb}$ and ρ are numerically equal in the two cases.

Again it is recognized from equation 34 that variations of feedback parameters only may result in some small differences between the ρ's in the two cases. Further adjustments of the resistive component of the load impedance, Z_L, could achieve the desired invariance exactly, but since such minor readjustments are seen to have negligible effects upon the output one can state that identities of $N_cI_c + N_fI_f$ and of $N_g - N_{gb}$ are sufficient to insure practical identity of instantaneous values of output current, i_L, in the two cases.

(C) As in the previous circuit it is seen that poorer rectifiers can be compensated for so as to obtain unchanged output with unchanged control, N_cI_c, if $\xi = (N_F/N_G)\sigma$ is made numerically equal in the two amplifiers, as stated in equation 3, and if invariance of N_G is observed as stated in equation 4.

(D) If, on the other hand, no readjustments of design are made, identity of H_0 yielding identity of outputs requires that N_cI_c be changed. Manipulations similar to those shown in D of the preceding section yield equation 5.

Full-Wave Circuit: Figure 1(C)

(A) Kirchhoff's voltage equations are

$$v_c = A N_c \left(\frac{dB_{\mathrm{I}}}{dt} + \frac{dB_{\mathrm{II}}}{dt} \right) + 2r_c i_c \quad (36)$$

$$v_g = \left(N_g \frac{dB_{\mathrm{I}}}{dt} + N_{gb} \frac{dB_{\mathrm{II}}}{dt} \right) + (r_g + r_{gb}) i_{\mathrm{I}} + v_{ba} + R_L(i_{\mathrm{I}} + i_{\mathrm{II}}) \quad (37)$$

$$v_g = -A \left(N_{gb} \frac{dB_{\mathrm{I}}}{dt} + N_g \frac{dB_{\mathrm{II}}}{dt} \right) - (r_g + r_{gb}) i_{\mathrm{II}} - v_{bc} - R_L(i_{\mathrm{I}} + i_{\mathrm{II}}) \quad (38)$$

With R_1 and R_2 the rectifier resistances at the instant considered

$v_{ba} = R_1 i_{\mathrm{I}}$ and $v_{bc} = R_2 i_{\mathrm{II}}$.

The following notation is introduced

$N_F = N_g + N_{gb}$, $N_G = N_g - N_{gb}$, and

$$i_F = \frac{i_{\mathrm{I}} + i_{\mathrm{II}}}{2} = {}^1/_2 i_L, \quad i_G = \frac{i_{\mathrm{I}} - i_{\mathrm{II}}}{2} \quad (39)$$

From equations 37 and 38 by subtraction

$$0 = A N_G \left(\frac{dB_{\mathrm{I}}}{dt} + \frac{dB_{\mathrm{II}}}{dt} \right) + v_{ba} + v_{bc} + 2(r_g + r_{gb} + 2R_L) i_F \quad (40)$$

which is quite similar to equation 30. In this case the presence of the load R_L (supposed purely resistive in this type of circuit) in the last term of equation 40 makes it even more plausible to neglect possibilities of commutation in the steady state with constant control voltage V_c and reasonably low control circuit impedance. For such conditions equation 17 is accepted. For $r_c \cong 0$ the reverse voltage across one rectifier is found as

$$v_r = -2(r_g + r_{gb} + 2R_L) i_F - R_d(i_F + |i_G|) \cong -(r_g + r_{gb} + 2R_L + R_d) i_L$$

The presence of R_L causes this expression to have far higher values than in the preceding cases, as is well known.

Integrating equation 40 over one half-cycle from an instant of transition of i_L through zero yields then for half-cycle averages the ratio

$$\frac{I_F}{I_G} = \frac{R_r - R_d}{R_r + R_d + 2r_g + 2r_{gb} + 4R_L} = \sigma \quad (41)$$

indicative of the reduction of average feedback action caused by less than perfect rectifiers.

(B) Moreover, by addition of equations 37 and 38

$$v_g = {}^1/_2 A N_G \left(\frac{dB_{\mathrm{I}}}{dt} - \frac{dB_{\mathrm{II}}}{dt} \right) + \rho i_G + {}^1/_2 A N_F \left(\frac{dB_{\mathrm{I}}}{dt} + \frac{dB_{\mathrm{II}}}{dt} \right) \quad (42)$$

where

$$\rho = \left[(r_g + r_{gb}) + \left\{ \frac{R_r(r_g + r_{gb} + 2R_L + 2R_d) + R_d(r_g + r_{gb} + 2R_L)}{R_r + R_d + 2r_g + 2r_{gb} + 4R_L} \right\} \right] \quad (43)$$

Equation 42 is formally identical to equation 33, while the magnetomotive force equations 28 and 29 apply. Hence equations 35 and 24 result, which indicate that identity of H's at corresponding times (and therefore identity of $i_G = (l/2N_G)(H_1 - H_2)$ obtains approximately if H_0, $N_G = N_g - N_{gb}$ and ρ are numerically equal in the two amplifiers under comparison. (It should be noted that identity of i_G here does not quite mean identity of output currents, i_L, since the rectifiers of the two amplifiers are supposed to have different leakages. Such leakages, however, while affecting greatly the control current requirements, do not cause appreciable differences in the output currents i_L that result in the two cases for identical i_G's.)

It should be pointed out that variations of feedback circuit parameters may result inherently in differences between the values of ρ for the two cases that may be somewhat larger than in the previous circuits. Some more care may be needed to judge whether invariance of ρ must be obtained, for example, by readjustments of R_L. In most practical cases, however, the condition of invariance of ρ still can be disregarded in terms of external characteristics.

(C) As in the previous cases it is seen that poorer rectifiers can be compensated for so as to obtain unchanged i_G (and therefore also practically unchanged output i_L) with unchanged control, N_cI_c, if invariance of $\xi = (N_F/N_G)\sigma$ and of $N_G = N_g - N_{gb}$ are obtained, as has been stated in equations 6 and 7.

(D) If, on the other hand, no readjustments of design are made, identity of H_0 yielding identities of i_G requires that N_cI_c be varied. Manipulations similar to those shown in (D) under the section "External Feedback Circuit" in this Appendix yield hereby equation 8.

References

1. RULES OF SIMILITUDE FOR MAGNETIC AMPLIFIER SYSTEMS, L. A. Finzi, H. L. Durand. *Proceedings of the National Electronic Conference* (Chicago, Ill.), 1951. (An abridgement of Technical Report Number 9 Contracts N7, United States Office of Naval Research 30306 and 30308.)

2. A THEORETICAL AND EXPERIMENTAL STUDY OF THE SERIES-CONNECTED MAGNETIC AMPLIFIER, H. M. Gale, P. D. Atkinson. *Proceedings*, Institution of Electrical Engineers (London, England), volume 96, part 1, 1949, pages 99–114.

3. A NEW THEORY OF THE MAGNETIC AMPLIFIER, A. G. Milnes. *Proceedings*, Institution of Electrical Engineers (London, England), volume 97, part 2, 1950, pages 460–74.

4. THE TRANSDUCTOR AMPLIFIER (book), U. Krabbe. Lindhska Bokteyskerict, Orebo, Sweden, 1947, page 69.

No Discussion

A Turns Index for Pulse Transformer Design

HAROLD W. LORD
FELLOW AIEE

FOR a good many years designers of power transformers and distribution transformers have had a guide number to use in starting their first design.[1,2] This is expressed by the equation $V_t = K\sqrt{\text{kva}}$ where V_t is the volts per turn to use in starting the first design; K is a constant which may vary between relatively narrow limits; and kva is the kilovolt-ampere rating of the transformer. In practice, the constant K is about 1.25 for shell-type power transformers, about 0.6 for core-type power transformers, and about 0.4 for core-type distribution transformers.

Experienced transformer designers may seldom use this guide since their experience, plus a standard practice book applicable to their production techniques and a file of proved designs, enables them to choose a core size or some other parameter at which to start the first design. However, the guide number is useful to those transformer designers who lack such experience and/or design information.

Pulse transformers are relatively new members of the transformer family. They originated during World War II when transformers were needed to match radar transmitter tubes to pulse modulators or to pulse cables. The pulses these transformers must handle are substantially rectangular in shape, the rise time and fall time each being approximately one-tenth of the nominal pulse width. The majority of radar systems employed pulse widths falling within the range of 0.5 microsecond to 3 microseconds. However, some systems had pulses as short as 0.1 microsecond while others had pulses as long as 20 microseconds. Pulse rates usually fell within the range of 60 pulses per second to 5,000 pulses per second. The duty cycle (product of pulse width and pulse rate) seldom exceeded 0.001.

Early designs of pulse transformers were determined by a combination of a small amount of design and a large amount of testing of developmental models in actual radar systems. By the end of World War II, the design methods had been improved to a point where the first model built usually tested to be satisfactory if it had been designed by an experienced pulse-transformer designer. The design information available[3,4] provided means for calculating the parameters a pulse transformer should have to perform satisfactorily. It also provided means for calculating the parameters of a given transformer design. However, the design procedure was principally one of cut and try. Even experienced pulse-transformer designers usually had to make several designs before obtaining one with suitable values of calculated parameters. Much of this cut-and-try procedure is due to the fact that they have no guide number upon which to base their initial design. Without such a guide the designer must rely wholly upon experience and upon extrapolation and interpolation from previous successful designs of other ratings. The wide range of pulse-width ratings of pulse transformers makes it difficult to choose a good starting point for the first design.

This paper will show how a guide number for pulse-transformer design may be derived through the combination of certain basic equations and with some simplifying assumptions. Then design procedures using the guide number will be outlined for the case of a resistance load and for the case of a biased diode or magnetron load.

Derivation of a Guide Number

Pulse transformers usually are constructed with single-layer primary and secondary windings. With such a configuration, the distributed capacitance of the secondary and the total equivalent leakage inductance referred to the secondary are both approximately proportional to the same coil dimension; namely, the spacing between the primary and secondary windings. Since the leakage inductance is approximately directly proportional to this spacing and the capacitance is inversely proportional to this spacing, then the product of the two may be considered a constant, thus making it possible to eliminate this spacing as a preliminary design parameter.

Successful pulse-transformer designs have ratios of core dimensions and coil dimensions which fall within relatively narrow limits. By assuming certain ratios between these core and coil dimensions which represent the average of usual practice, it becomes possible to generalize the design relationships to such an extent that all of the coil and core dimensions are eliminated from preliminary considerations. The only parameter left is that of the turns of the windings. By referring all values to the secondary winding, the secondary turns can be made a function of the performance requirements and properties of the dielectric and magnetic materials available for use in pulse transformers.

Nomenclature

The following nomenclature and units will be used throughout this paper:

Paper 52-125, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE North Eastern District Meeting, Binghamton, N. Y., April 30–May 2, 1952. Manuscript submitted March 26, 1951; made available for printing February 28, 1952.

Harold W. Lord is with the General Electric Research Laboratory, Schenectady, N. Y.

The author acknowledges and appreciates the assistance of P. F. Fenoglio in carrying out the validity checks reported in this paper. Part of the work forming the basis for this paper was supported by contracts with the Signal Corps Engineering Laboratories, Fort Monmouth, N. J.

C_L = total load capacitance including transformer bushings and all stray wiring capacitances, micromicrofarads
C_d' = distributed capacitance of secondary, micromicrofarads
L_1' = total leakage inductance referred to secondary, microhenrys
L_2' = primary exciting inductance referred to secondary, henrys
R_L = equivalent resistance (static impedance) of load on transformer secondary, ohms
R_L' = dynamic resistance of load on transformer secondary, ohms
R_p' = internal resistance of pulse generator as referred to secondary, ohms
R_e = equivalent core loss resistance as referred to secondary, ohms
N_s = total of secondary turns in series connection
E_o = output (secondary) pulse voltage, volts
E_d = output (secondary) pulse-voltage droop, volts
E_L' = diode bias voltage or magnetron firing voltage
V_t = volts per turn $= E_o/N_s$
I_o = output pulse current, amperes
I' = instantaneous magnetizing current as referred to secondary, amperes
I_m' = primary peak magnetizing current at end of pulse as referred to the secondary, amperes
T_o = pulse width, microseconds
T_r = rise time of pulse, microseconds
ΔB = flux density change during pulse, lines per square inch of gross core area
μ_p = pulse permeability of core with a flux-density change of ΔB in time T_o
a_i = width of core leg, inches
h = stack height of core, inches
A_i = gross core area $= a_i h$, square inches
l_w = length of core window, inches
a_w = width of core window, inches
l_i = mean length of magnetic circuit, inches
t = wire traverse of each winding layer, inches
s = insulation thickness between primary and secondary, inches
k = dielectric constant of insulation between primary and secondary
d_1 = thickness (radial build) of primary, inches
d_2 = thickness (radial build) of secondary, inches
l_c = mean length of turn of coil, inches
r = ratio of secondary turns to primary turns

GENERAL ASSUMPTIONS

The following general assumptions are made

1. Primary capacitance negligibly small
2. Primary-to-secondary capacitance may be lumped on secondary or it is negligibly small
3. Step-up ratio r is more than 3
4. $a_i = a_w = l_w/3$
5. $h = 1.5a_i (A_i = 1.5a_i^2)$
6. $t = 0.6l_w$
7. d_1 and $d_2 << s$
8. $l_i = 4l_w$
9. $l_c = 7.5a_i = 2.5l_w = 0.625l_i$
10. $R_e >> R_L$
11. $L_2'/R_L >> T_o$

BASIC EQUATIONS

The following basic equations are used:

Pulse-transformer voltage equation

$$E_o = \frac{N_s \Delta B A_i 10^{-8}}{T_o 10^{-6}}$$

$$A_i = \frac{E_o T_o 10^2}{N_s \Delta B} \text{ square inches} \tag{1}$$

Magnetizing inductance referred to secondary

$$L_2' = \frac{3.2N_s^2 10^{-8}}{\dfrac{l_i}{\mu_p A_i}} = \frac{3.2N_s^2 \mu_p A_i 10^{-8}}{l_i} \text{ henrys} \tag{2}$$

Since R_e is assumed much greater than R_L and L_2'/R_L is assumed much greater than T_o, then peak magnetizing current is related to E_o as follows

$$E_o = L_2' \frac{\Delta I'}{\Delta T} = \frac{L_2' I_m'}{T_o 10^{-6}} \tag{3}$$

This assumes I' varies as a linear function of time instead of as an exponential function of time. This does not introduce appreciable errors if $L_2'/R_L >> T_o$.

If the transformer is assumed to have only a single coil:
Total leakage inductance referred to the secondary

$$L_1' = 3.2N_s^2 \frac{l_c}{t}\left(s + \frac{d_1 + d_2}{3}\right) 10^{-2} \mu \text{ henrys} \tag{4}$$

The distributed capacitance is then one-third the capacitance as calculated on the basis of the primary and secondary being considered equipotential surfaces s distance apart. Hence

$$C_d' = \frac{1}{3} \frac{0.225ktl_c}{s} = \frac{0.075ktl_c}{s} \text{ micromicrofarads} \tag{5}$$

The guide number is derived as follows:

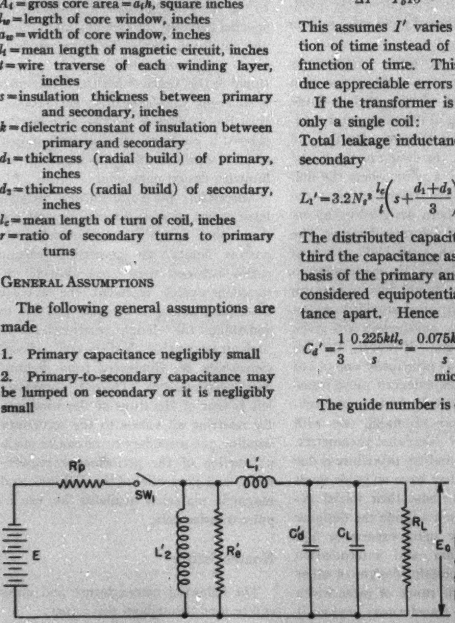


Figure 1. Equivalent circuit of power source, pulse transformer, and resistive load

Combining equations 1 and 2 to eliminate A_i

$$L_2' = \frac{3.2N_s^2 \mu_p E_o T_o 10^{-6}}{l_i N_s \Delta B} = \frac{3.2N_s \mu_p E_o T_o 10^{-6}}{l_i \Delta B} \tag{6}$$

Combining equations 3 and 6 to eliminate E_o, L_2', and T_o

$$L_2' = \frac{3.2N_s \mu_p L_2' I_m' T_o 10^{-6}}{l_i \Delta B T_o 10^{-6}}$$

$$l_i = \frac{3.2N_s \mu_p I_m'}{\Delta B} \tag{7}$$

Since d_1 and d_2 are assumed $<< s$ equation 4 may be written

$$L_1' = 3.2N_s^2 \frac{l_c}{t}(s) 10^{-2} \text{ microhenrys} \tag{8}$$

Multiplying equations 5 and 8, and extracting the square root

$$L_1' C_d' = 0.24N_s^2 l_c^2 k 10^{-2}$$

$$\sqrt{L_1' C_d'} = N_s l_c \sqrt{0.24k}\, 10^{-1}$$

$$l_c = \frac{10\sqrt{L_1' C_d'}}{N_s \sqrt{0.24k}} = \frac{20.3\sqrt{L_1' C_d'}}{N_s \sqrt{k}} \tag{9}$$

Since l_c is assumed equal to 0.625 l_i, then

$$0.625 l_i = \frac{20.3\sqrt{L_1' C_d'}}{N_s \sqrt{k}};$$

$$l_i = \frac{3.2\sqrt{L_1' C_d'}\, 10}{N_s \sqrt{k}} \tag{10}$$

Combining equations 10 and 7 to eliminate l_i

$$\frac{3.2N_s \mu_p I_m'}{\Delta B} = \frac{3.2\sqrt{L_1' C_d'}\, 10}{N_s \sqrt{k}}$$

$$N_s^2 = \frac{\Delta B \sqrt{L_1' C_d'}\, 10}{\mu_p I_m' \sqrt{k}}$$

$$N_s = \sqrt{\frac{10 \Delta B \sqrt{L_1' C_d'}}{\mu_p I_m' \sqrt{k}}} \tag{11}$$

We will call this relationship the turns index since it enables the designer to determine the number of secondary turns to use for his first design. This is possible since all parameters in the right-hand side of the equation are either properties of the core and insulating materials available or fixed by the required performance.

If the transformer has two coils of the same relative dimensions, one on each leg and with the secondary windings in series, the following additional assumptions are made

12. Insulation thickness of s_2 of coil having higher secondary voltage to ground and primary is three times the spacing s_1 of the other coil ($s_2 = 3s_1$). If s_1 is made equal to $s/2$, $s_2 = 1.5s$, then the effective distributed

capacitance of such a combination is twice that of the single-coil transformer.

13. The secondary turns on each leg are $N_s/2$, hence the leakage inductance will be one-half that of the single-coil type.

The $L_1'C_d'$ product of this transformer type therefore will be the same as that of the single-layer type, hence the same turns index will hold as a guide for both types. The advantage of the 2-coil construction lies in the fact that the total available wire traverse is twice that of the single-layer type on a given core; therefore, it will have higher power-handling capabilities than that of the corresponding single-layer type.

Brief Description of Design Procedure Using the Turns Index

For Resistance Loads

The entire pulse generator, pulse transformer, and load (reduced to unity turns ratio) may be represented by the equivalent circuit of Figure 1.

Since the frequency band-pass characteristics of most pulse transformers are reasonably wide, this circuit can be reduced to three simplified circuits, one effective during the leading edge of the pulse, a second for the top of the pulse, and a third which applies for the trailing edge of the pulse. These are shown in Figures 2(A), 2(B), and 2(C) respectively.

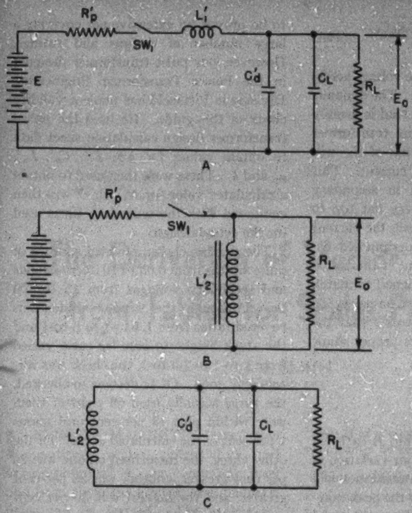


Figure 2 (left). Simplified circuits applicable to different portions of pulse

A. To leading edge
B. To top
C. To trailing edge

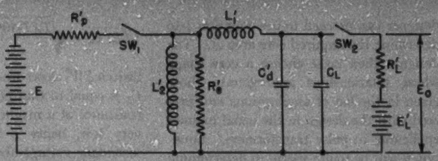


Figure 3 (above). Equivalent circuit of power source, pulse transformer, and magnetron or biased diode load

Using Figure 2(A) for the leading edge of the pulse, one can determine the value of the quantity $\sqrt{L_1'C_d'}$. On page 222 of "Electronic Transformers and Circuits"[3] a graph shows the wave shapes for this circuit. Since it is recommended that $\sqrt{L_1'/C_d'+C_L}10^3=R_L$ for minimum rise time with tolerable overshoot, page 536 "Pulse Generators,"[4] and since $R_L=R_p'$ for best matching of a line-type pulse generator, then the curve $k=1$ (Lee's nomenclature[3]) should be used. This gives a rise time to 90 per cent of final voltage which is 63 per cent of the resonant period of the transformer and load.

Hence

$$T_r=0.63\ [2\pi\sqrt{L_1'(C_d'+C_L)}]\times10^{-3}$$
$$=4\sqrt{L_1'(C_d'+C_L)}\times10^{-3} \qquad (12)$$

and for best shape of leading edge

$$\sqrt{\frac{L_1'}{C_d'+C_L}}10^3=R_L \qquad (13)$$

Equations 12 and 13 may be combined to give

$$L_1'=T_rR_L/4 \qquad (14)$$

$$(C_d'+C_L)=\frac{T_r\times10^6}{4R_L},\ C_d'=\frac{T_r\times10^6}{4R_L}-C_L \qquad (15)$$

Multiplying equations 14 and 15 and taking the square root of the product

$$\sqrt{L_1'C_d'}=\sqrt{\frac{T_r(T_r10^6-4R_LC_L)}{4}} \qquad (16)$$

Equation 16 therefore provides a means for obtaining a value for the square root of the $L_1'C_d'$ product in terms of the load resistance (R_L), shunt capacitance (C_L), and the desired rise time (T_r). Equation 14 determines the required value of leakage inductance in terms of the load resistance and the desired rise time.

It should be noted that if C_d' is shown by equation 15 to be negative in order to meet the desired rise time, then either the load shunt capacitance must be reduced or the rise time allowed to be longer.

Using Figure 2(B) for the top of the pulse, one can determine a value of I_m' which will provide proper flatness across the top of the pulse. The specifications should call for an allowable maximum value of droop across the top of the pulse of some percentage of the nominal output pulse voltage E_o. This droop in voltage (E_d = per cent E_o) will be proportional to the peak magnetizing current I_m'. If the load is matched to the internal resistance of the pulse generator, then the allowable peak magnetizing current I_m' will be the load current multiplied by twice the allowable per cent droop, hence one may write

$$I_m'=\frac{2E_d}{E_o}\frac{E_o}{R_L}=\frac{2E_d}{R_L} \qquad (17)$$

Equation 17 then provides a means for determining a value for I_m' in terms of the allowable droop voltage and the load resistance. If the pulse generator has a lower or higher internal resistance, then the expression for maximum allowable I_m' becomes

$$I_m'=E_d\frac{(R_L+R_p')}{R_LR_p'} \qquad (17A)$$

The trailing edge shape may be determined by information given on pages 226–30 of "Pulse Generators."[4] The value for L_2' may be obtained by using equation 3 and the value of I_m' as determined in equation 17(A). However, the exact determination of the trailing edge may well be left until after the design has been determined more accurately.

The designer has now determined all of the parameters of equation 11 which depend upon the required performance. The rest of the values to be used in equation 11 are properties of the core material and insulating materials.

Having determined a value for the turns index N_s by using equation 11, the designer can proceed to use equation 1 to determine the required core area A_i. The designer then must choose a core shape whose relationships approximate those given in the list of assumptions and proceed with the design in the usual manner applicable to pulse transformers. Equation 8 will be found useful in determining a spacing s and traverse t which will provide the required value of L_1'. For a 2-coil type, the average spacing should be double that as determined upon the basis of all the secondary considered as a single layer on one leg of the core.

After completing the first design, if the performance is inferior to that required, then the secondary turns may have to be reduced and another design carried thru. This should be continued until either a satisfactory design is obtained or until the designer is convinced that the requirements cannot be met. Such may be the case when the required value of C_d' is only a few micromicrofarads.

For Magnetron and Biased Diode Loads

The equivalent circuit for this type of load is shown in Figure 3.

Switch SW_1 represents the modulator tube and closes while SW_2 remains open until the voltage across capacitor C_L equals the voltage E_L'. During this initial rise period the transformer is, therefore, open-circuited and the voltage rises much more rapidly than for the resistive load case. Under these circumstances and with

$$10^3\sqrt{\frac{L_1'}{C_d'+C_L}}=R_L=\frac{E_o}{I_o}$$

the voltage rises to 90 per cent of the final value in approximately 25 per cent of the resonant period of the transformer and load; see page 565 of "Pulse Generators."[4] Equation 12 then becomes

$$\begin{aligned}T_r&=0.25[2\pi\sqrt{L_1'(C_d'+C_L)}10^{-3}]\\&=1.6\sqrt{L_1'(C_d'+C_L)}10^{-3}\qquad\text{(12A)}\end{aligned}$$

Equation 14 becomes

$$L_1'=T_rR_L/4\qquad\text{(14A)}$$

Equation 15 becomes

$$C_d'=\frac{T_r10^6}{1.6R_L}-C_L\qquad\text{(15A)}$$

Equation 16 becomes

$$\sqrt{L_1'C_d'}=\sqrt{\frac{T_r(T_r10^6-1.6R_LC_L)}{1.6}}\qquad\text{(16A)}$$

Switch SW_2 closes when the voltage across C_L is equal to that of E_L'. The dynamic resistance of a magnetron load is usually very low, hence a rise in transformer secondary voltage above that of E_L' will cause a large flow of load current. This effectively stops the rise in secondary voltage and produces a very flat top to the voltage pulse. However, the current pulse may have large superimposed oscillations or a long rise time if the transformer is not properly designed to match the load. It has been shown on pages 566 and 567 of "Pulse Generators"[4] that the current to the load has the proper shape when

$$10^3\sqrt{\frac{L_1'}{C_d'+C_L}}=R_L$$

Assuming this relationship is satisfied by the design, then the major variation in load current due to the transformer will be the current droop due to the peak magnetizing current I_m'. Since the load voltage is essentially constant, the pulse generator supplies a relatively constant current to the transformer and load. The amperes of current droop in the load will then be very nearly equal to the peak magnetizing current I_m'. The designer can, therefore, use the allowable current droop in the load as a value for I_m' in equation 11.

Switch SW_2 is opened as soon as the voltage E_o falls below E_L' at the start of the trailing edge of the pulse. The load is, therefore, disconnected from the transformer at this time and only the core losses are effective in damping the oscillations and back swing. Since R_e' is much higher than R_L, the oscillations and backswing peaks will be much higher for this case. The same reference given for the loaded case includes information applicable to this case.

Except for the modifications previously indicated, the pulse-transformer design procedure for this case is similar to that for the resistive load case.

Check on Validity

This pulse-transformer turns index is a fairly recent development and so has yet to be proved by extensive use covering a large number of designs and ratings. However, our pulse-transformer designer in the Power Transformer Engineering Division in Pittsfield has made a validity check of this guide. He used the pulse-transformer design calculation sheet data to obtain values for ΔB, L_1', C_d', I_m', μ_p and k. These were then used to obtain a calculated value for N_s, which was then compared with the secondary turns used for the actual design.

The designs chosen covered a range of pulse widths from 0.05 to 10 microseconds and secondary voltages from 15 to 100 kv. Although the window dimensions covered ratios from 1.4-to-1 to 6-to-1 and the stack height to core-leg ratio varied from 1-to-1 to 2.3-to-1, the check was surprisingly good. Of 11 designs so checked, the turns actually used on eight of them were within plus 18 per cent and minus 1 per cent of the calculated value. Of the other three, the turns used on one was 26 per cent greater, another was 33 per cent greater, and the highest was 50 per cent greater than calculated values. The latter was a 100-kv unit with a window dimension ratio of 4-to-1. Six of the designs were within plus 11 per cent and minus 1 per cent of the calculated value. The average of all 11 units exceeded the calculated average value by 15 per cent.

These data indicate that this turns index provides a usable value even when the actual dimension ratios depart to some extent from those upon which the guide is based. It is probably even superior to most educated guesses by an experienced pulse-transformer designer. Such a guide should, therefore, be a useful tool in the hands of both inexperienced and experienced pulse-transformer designers.

References

1. Electrical Machine Design (book), **Alexander Gray.** McGraw-Hill Book Company, Inc., New York, N. Y., first edition, 1913.

2. The Performance and Design of Alternating Current Machines (book), **M. G. Say, E. N. Pink.** Pitman Publishing Corporation, New York, N. Y., 1936.

3. Electronic Transformers and Circuits (book), **Reuben Lee.** John Wiley and Sons, Inc., New York, N. Y., 1947.

4. Pulse Generators (book), **G. N. Glasoe, J. V. Levacqz.** Volume 5 of Radiation Laboratory Series, first edition. McGraw-Hill Book Company, Inc., New York, N. Y., 1948.

No Discussion

The Derivation of Vector Potential from Tables for Scalar Potential

J. J. SMITH
MEMBER AIEE

IN ELECTROMAGNETIC theory two quantities, scalar potential V and vector potential A, are commonly used and from them the electric and magnetic fields can be derived. The author recently published tables of Green's Functions[1] for the solution of partial differential equations in rectangular co-ordinates. The Green's Functions as given in those tables only include scalar potentials. However, in dealing with problems involving fields due to currents or permanent magnets vector potentials are used. These also can be derived from the tables of Green's Functions when required. Since this was not included in the tables on account of their length, it will be covered here.

An examination of the literature shows that vector potential generally is defined in two different ways and, to apply the results given in the tables, it is necessary to discuss these definitions and add a third definition which allows a comparison between them to be made. For this reason these definitions of vector potentials will be given first and examples of their use given in later sections of the paper.

Definitions of Vector Potential

DEFINITION I

The following is that given in the American Standards Association Definitions of Electrical Terms:[2]

> "A vector potential field of a tubular vector field is a vector field which has the property that its curl is the tubular vector field.
>
> "Example: The vector potential field of the current density in an electric circuit is the field of magnetic intensity in the region occupied by the current.
>
> "Note: Any vector field can be expressed as the gradient of a scalar potential field plus the curl of a vector potential field, either of which may be zero."

This definition has given rise to some confusion since it is too general. Under it in free space we might have a vector potential A_1 such that curl $A_1=B$ and also a vector potential A_2 such that curl $A_2=H$. Since $B=\mu H$, then $A_1=\mu A_2$, and unless it is specifically stated that A is related to B or H, confusion may arise. There is no agreement on this in the literature and so it is necessary to make a choice. We shall use the equation μ curl $A=B$ and this notation will be followed in the remainder of this paper.

DEFINITION II

This definition is based upon Lorentz's solution of Maxwell's equation which states that the potential is given by

$$V=\int\int\int\frac{[\rho]dx_1dy_1dz_1}{R} \tag{1}$$

and the vector potential is given by

$$A=\int\int\int\frac{[\rho u]dx_1dy_1dz_1}{R} \tag{2}$$

where[3] R is the distance of the field point, that is, of the point x,y,z of the field where the potentials are required, from the position x_1,y_1,z_1 of the volume element $dx_1dy_1dz_1$ at the time t. The volume integration is extended over all of space. The brackets [...] mean that the values of ρ and u are to be taken at the point $t-R/c$ and not for the actual time t; that is, the value of ρ and u correspond to the time of emission of a disturbance which leaves the element $dx_1dy_1dz_1$ at (x_1,y_1,z_1) so as to reach the field point at the time t.

This definition is more specific than Definition I. The flux density B is obtained from it by the equation μ curl $A=B$. Also using equations 1 and 2 the force on a unit charge is $E_s=\partial V_s/\partial s+\mu\partial A_s/\partial t$.

DEFINITION III

The vector potential A at a point may be defined as follows in rectangular coordinates. From the point x,y,z draw a line to infinity in the direction of one of the axes. Then the total flux in the direction of one of the other axes crossing this line between the point and infinity may be taken as μ times the component of vector potential in the direction of the third axis at the point. Since there are two such axes, this gives two components of vector potential. The third component of vector potential in the direction of the first line drawn to infinity is zero. μ is the permeability of the medium. This definition is shown graphically in Figure 1.

Discussion of Definition III. Definition III has the advantage that it allows us to develop the idea of vector potential along lines customarily used in connection with the developing of scalar potential at a point as the work done in moving a unit charge from a point at zero potential, generally infinity, up to the point in the case where there is no variation with time as shown below.

Take a unit current in a unit length of wire oriented in the direction of the x axis and move it along a path parallel to the y axis keeping its orientation fixed. Then the work done in moving it from an infinite distance to the point (x,y,z) is μ times the x component of vector potential A_x at the point (x,y,z) where μ is the permeability of the medium. Similarly if the unit current in unit length of wire be oriented in the y direction, then the work done in moving it parallel to the y axis from an infinite distance to the point (x,y,z) is μ times the y component of vector potential A_y at the point which for this case is obviously zero. Finally, if the unit

---

Paper 52-129, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE North Eastern District Meeting, Binghamton, N. Y., April 30–May 2, 1952. Manuscript submitted October 22, 1951; made available for printing February 28, 1952.

J. J. SMITH is with the General Electric Company, Schenectady, N. Y.

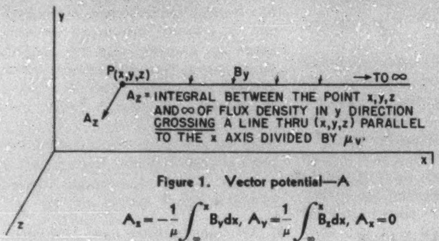

y
P(x,y,z)
By
→TO ∞
A_z = INTEGRAL BETWEEN THE POINT x,y,z AND ∞ OF FLUX DENSITY IN y DIRECTION CROSSING A LINE THRU (x,y,z) PARALLEL TO THE x AXIS DIVIDED BY μ_v.
A_z
x
z

Figure 1. Vector potential—A

$$A_z=-\frac{1}{\mu}\int_{\infty}^{x}B_y dx,\ A_y=\frac{1}{\mu}\int_{\infty}^{x}B_z dx,\ A_x=0$$

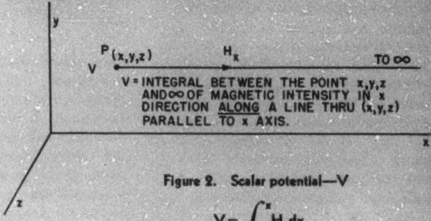

y
P(x,y,z)
V
H_x
TO ∞
V = INTEGRAL BETWEEN THE POINT x,y,z AND ∞ OF MAGNETIC INTENSITY IN x DIRECTION ALONG A LINE THRU (x,y,z) PARALLEL TO x AXIS.
x
z

Figure 2. Scalar potential—V

$$V=\int_{\infty}^{x}H_x dx$$

current in unit length of wire be oriented in the z direction, then the work done in moving it parallel to the y axis from a point at an infinite distance to the point (x,y,z) is μ times the z component of vector potential at the point. These definitions can be written

$$A_{2x}=-\frac{1}{\mu}\int_{\infty}^{y}B_z dy \qquad A_{2y}=0$$

$$A_{2z}=\frac{1}{\mu}\int_{\infty}^{y}B_x dy \quad (3)$$

In equation 3, the unit currents oriented in the directions of the three axes were moved up along a line parallel to the y axis. Since they also can be moved up along a line parallel to the z axis or the x axis, two other sets of equations can be found which can be taken equally well as the components of vector potential as follows.

If the unit current oriented successively in the direction of the three axes be moved along a line parallel to the z axis, the following can be taken as the components of vector potential

$$A_{3x}=\frac{1}{\mu}\int_{\infty}^{z}B_y dz \qquad A_{3y}=-\frac{1}{\mu}\int_{\infty}^{z}B_x dz$$

$$A_{3z}=0 \quad (4)$$

If the unit currents oriented successively in the direction of the three axes be moved along a line parallel to the x axis, the following can be taken as the components of vector potential

$$A_{1x}=0 \qquad A_{1y}=\frac{1}{\mu}\int_{\infty}^{x}B_z dx$$

$$A_{1z}=-\frac{1}{\mu}\int_{\infty}^{x}B_y dx \quad (5)$$

It will be found that the curl of the different expressions for vector potential in equation 3, 4 or 5 will give the same value for the flux densities B_x, B_y, and B_z. The difference between equations 3, 4, and 5 is due to the fact that these definitions do not define a unique value for the vector potential. To any of these equations may be added a quantity whose curl is zero. The gradient of a scalar is such a quantity. As an example, use may be made of such a quantity to turn the vector potential given by equation 3 into that given by equation 5. To do this add

$$\int_{\infty}^{y}\frac{\partial A_{1y}}{\partial x}dy \text{ to } A_{2x},\ \int_{\infty}^{y}\frac{\partial A_{1y}}{\partial y}dy \text{ to } A_{2y},$$

$$\text{and } \int_{\infty}^{y}\frac{\partial A_{1y}}{\partial z}dy \text{ to } A_{1z}$$

and obtain

$$A_{2x1}=-\frac{1}{\mu}\int_{\infty}^{y}B_z dy+\int_{\infty}^{y}\frac{\partial A_{1y}}{\partial x}dy=0$$

$$A_{2y1}=\int_{\infty}^{y}\frac{\partial A_{1y}}{\partial y}dy=A_{1y} \quad (6)$$

$$A_{2z1}=\frac{1}{\mu}\int_{\infty}^{y}B_x dy+\int_{\infty}^{y}\frac{\partial A_{1y}}{\partial z}dy=A_{1z}$$

By a similar procedure the vector potential in equation 4 may be obtained from equation 3, confirming the fact that the curl of either equation 3, 4, or 5 gives the same values of B_x, B_y, and B_z. It will be noted also that by this procedure one component of vector potential always can be made equal to zero without loss of generality.

The usual symmetrical forms of vector potential given in textbooks may be obtained by adding equations 3, 4, and 5 to obtain

$$A_x=\frac{1}{3}(A_{1x}+A_{2x}+A_{3x})$$

$$A_y=\frac{1}{3}(A_{1y}+A_{2y}+A_{3y}) \quad (7)$$

$$A_z=\frac{1}{3}(A_{1z}+A_{2z}+A_{3z})$$

and it is readily checked that curl $A=B$.

Relations Between Scalar Potential and Vector Potential When There Is No Variation with Time

The use of Definition III makes the definition of vector potential analogous to that of scalar potential. The scalar potential may be defined as the work done in bringing up a unit charge or pole from infinity, assumed to be zero potential, to the point giving for a unit pole

$$V_1=\frac{1}{\mu}\int_{\infty}^{x}B_x dx \text{ or } V_2=\frac{1}{\mu}\int_{\infty}^{y}B_y dy \text{ or}$$

$$V_3=\frac{1}{\mu}\int_{\infty}^{z}B_z dz \quad (8)$$

the μ being introduced since by definition $V=\int H_s ds=1/\mu\int B_s ds$.

This definition is shown graphically in Figure 2 and it is interesting to compare it with Figure 1. In the case of bringing up a charge, there is no effect due to the orientation of the charge. In the case of bringing up a unit current from infinity, the work done depends on the orientation of the current and this is taken into account by taking the three separate cases where the unit current is oriented in the x, y, and z directions and calling these the x, y, and z components of vector potential.

It will be noted that three different expressions are given for vector potential in equations 3, 4, and 5. Similarly it will be noted that three different expressions are given for scalar potential in equation 8. In the latter case V_1, V_2, and V_3 may differ by constants k_1, k_2, and k_3 and still give the same value of B_x, B_y, and B_z since $dk_1/dx=0$, and so forth. Thus the value of the scalar potential is not necessarily unique. Also the scalar potential in the magnetic case is many valued since it depends on the current enclosed by or the sheet of magnetic dipoles crossed by any closed path. The latter case is avoided by recognizing that the sheet of dipoles cannot be crossed. Similarly the value of vector potential is not unique and in equations 3, 4, and 5 we have three sets of vector potentials which differ by a quantity whose curl is zero.

If now equation 8 is written for the case when there is no variation in time so that $\partial A/\partial t=0$ in the form where $V=V_1$, V_2, or V_3

$$\frac{\partial V}{\partial x}=-B_x \quad \frac{\partial V}{\partial y}=-B_y \quad \frac{\partial V}{\partial z}=-B_z \qquad (9)$$

then it is apparent that on substituting these values for B_x, B_y, and B_z in equation 3

$$A_{2x}=\int_\infty^y \frac{\partial V}{\partial z}dy,\ A_{2y}=0, \text{ and}$$

$$A_{2z}=-\int_\infty^y \frac{\partial V}{\partial x}dy \qquad (10)$$

Similar expressions can be obtained from equations 4 and 5. Thus when there is no variation with time, the vector potential can be derived from the scalar potential by these or similar equations derived from equations 4 and 5.

The converse also holds. Write equation 3 in the form

$$B_x=\frac{\partial A_{2z}}{\partial y} \qquad B_y=\frac{\partial A_{2x}}{\partial z}-\frac{\partial A_{2z}}{\partial x}$$

$$B_z=-\frac{\partial A_{2x}}{\partial y} \qquad (11)$$

then equation 8 can be written

$$V_1=\int_\infty^x \frac{\partial A_{2z}}{\partial y}dx$$

$$V_2=\int_\infty^y\left(\frac{\partial A_{2x}}{\partial z}-\frac{\partial A_{2z}}{\partial x}\right)dy \qquad (12)$$

$$V_3=-\int_\infty^z \frac{\partial A_{2x}}{\partial y}dz$$

so that when there is no variation with time if the vector potential is known, the scalar potential can be derived from it by these three equations or similar ones derived from equations 4 and 5. All these equations may give the same form for V or possibly differ by a constant.

The Relation Between the Vector Potentials Obtained from Definitions II and III

Let us now examine the relation between the expression for vector potential given under Definition II by equation 2 and that given under Definition III by equations 3, 4, and 5. Since from equation 2 a current flowing in the x direction gives a vector potential whose components are A_x, 0, and 0 we have

$$B_{xx}=0 \quad B_{yx}=\mu\frac{\partial A_x}{\partial z} \quad B_{zx}=-\mu\frac{\partial A_x}{\partial y} \qquad (13)$$

where the second subscript denotes it is the flux produced by current in the x direction only. Similarly for current in the y direction the vector potential has the components 0, A_y, 0 and

$$B_{xy}=-\mu\frac{\partial A_y}{\partial z} \quad B_{yy}=0 \quad B_{zy}=\mu\frac{\partial A_y}{\partial x} \qquad (14)$$

and for currents in the z direction

$$B_{xz}=\mu\frac{\partial A_z}{\partial y} \quad B_{yz}=-\mu\frac{\partial A_z}{\partial x} \quad B_{zz}=0 \qquad (15)$$

From these, if B_x,B_y, and B_z are the total flux density in the x, y, and z directions, respectively

$$B_x=B_{xy}+B_{xz},\ B_y=B_{yx}+B_{yz}, \text{ and}$$

$$B_z=B_{zx}+B_{zy} \qquad (16)$$

Substitute these values of B_x, B_y, and B_z in equation 3 to obtain

$$A_{2z}=\frac{1}{\mu}\int_\infty^y (B_{xy}+B_{xz})dy$$

$$=\int_\infty^y\left(-\frac{\partial A_y}{\partial z}+\frac{\partial A_z}{\partial y}\right)dy$$

$$=A_z-\int_\infty^y \frac{\partial A_y}{\partial z}dy$$

$$A_{2y}=0 \qquad (17)$$

$$A_{2x}=-\frac{1}{\mu}\int_\infty^y (B_{zx}+B_{zy})dy$$

$$=-\int_\infty^y\left(-\frac{\partial A_x}{\partial y}+\frac{\partial A_y}{\partial x}\right)dy$$

$$=A_x-\int_\infty^y \frac{\partial A_y}{\partial x}dy$$

Thus the vector potential whose components are A_{2x}, A_{2y}, and A_{2z} derived from Definition III differs from the vector potential whose components are A_x, A_y, and A_z derived from Definition II by the integral with respect to y of the gradient of A_y. Since the curl of the gradient of A_y is equal to zero the two vectors give the same values of magnetic flux density B. Hence for steady-state conditions, although the definition of vector potential given in equation 2 may lead to a different value of the vector potential than that defined in equation 3, 4, or 5; both definitions lead to the same values of magnetic flux density B.

The Relation Between Vector Potential in Definition I and in Definitions II and III

The definition of vector potential in Definition II given in equation 2 only defines the vector potential A. It does not give the rules for obtaining the flux density, say, B from the vector potential. It was supplemented by the equation μ curl $A=B$ which is equivalent to the statement in Definition I. Therefore, Definition I might be combined with Definition II to obtain a definition of vector potential and the way to use it to obtain the flux densities.

Definition I may be obtained by noting that in any rectangular parallelepiped the total flux ϕ which emerges from the surface is zero in the case of a tubular field (no charges inside the surface) and thus

$$\phi=\int\int B_x dydz+\int\int B_y dzdx+\int\int B_z dxdy \qquad (18)$$

Since $\phi=0$ differentiating with respect to x, y, and z successively gives

$$0=\frac{\partial B_x}{\partial x}+\frac{\partial B_y}{\partial y}+\frac{\partial B_z}{\partial z}=\text{div } B \qquad (18A)$$

The equation for ϕ, however, can be used in another way. Let us assume that the rectangular parallelepiped extends from the point x,y,z to infinity in the x direction and that in the y and z direction its dimensions remain finite. Then equation 18 can be written when $\phi=0$ and assuming the flux density at infinity tends to zero

$$\int\int B_x dydz=-\int\int_\infty^x B_y dz-\int\int_\infty^x B_z dydx$$

Differentiate with respect to y and z obtaining

$$B_x=-\frac{\partial}{\partial y}\int_\infty^x B_y dx-\frac{\partial}{\partial z}\int_\infty^x B_z dx$$

Similarly

$$B_y=-\frac{\partial}{\partial z}\int_\infty^y B_z dy-\frac{\partial}{\partial x}\int_\infty^y B_x dy \qquad (19)$$

and

$$B_z=-\frac{\partial}{\partial x}\int_\infty^z B_x dz-\frac{\partial}{\partial y}\int_\infty^z B_y dz$$

which may be checked with equation 7.

The Scalar and Vector Potential Due to a Point Source

As a simple example to illustrate this, consider a point source (charge or magnetic pole) in free space. The scalar potential V due to it is

$$V=1/R \qquad (20)$$

where $R^2=x^2+y^2+z^2$. The flux densities B_x, B_y, B_z are the gradients of V multiplied by μ so that

$$B_x=\mu\frac{\partial}{\partial x}\left(\frac{1}{R}\right) \text{ and so forth} \qquad (21)$$

The vector potential for the point source is found from equation 3 to be

$$A_{2x}=-\int_\infty^y \frac{\partial}{\partial z}\left(\frac{1}{R}\right)dy \qquad A_{2y}=0$$

$$A_{2z}=\int_\infty^y \frac{\partial}{\partial x}\left(\frac{1}{R}\right)dy \qquad (22)$$

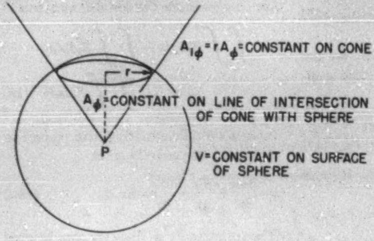

$A_{1\phi}=rA_\phi$=CONSTANT ON CONE
r
A_ϕ=CONSTANT ON LINE OF INTERSECTION OF CONE WITH SPHERE
P
V=CONSTANT ON SURFACE OF SPHERE

Figure 3. Scalar and vector equipotentials for a point charge at P

or

$$A_{2x}=-\frac{z}{R(R+y)} \qquad A_{2y}=0$$
$$A_{2z}=\frac{x}{R(R+y)} \qquad (23)$$

However, if equation 4 had been used, then it would be found that

$$A_{3x}=\frac{y}{R(R+z)} \qquad A_{3y}=-\frac{x}{R(R+z)}$$
$$A_{3z}=0 \qquad (24)$$

and similarly if equation 5 were used it would be found that

$$A_{1x}=0 \qquad A_{1y}=\frac{z}{R(R+x)}$$
$$A_{1z}=-\frac{y}{R(R+x)} \qquad (25)$$

It can be verified that any of these sets of vector potential lead to the same value of B_x, B_y, and B_z. Thus we have in equations 23, 24, and 25 three different sets of vector potentials which can be used to derive the field due to a point source in space. If cylindrical co-ordinates r, x, and ϕ are used, then it is shown readily that the values of A_{1y} and A_{1z} in equation 25 correspond to

$$A_r=0 \text{ and } A_\phi=\frac{(y^2+z^2)^{1/2}}{R(R+x)} \text{ and again}$$
$$A_{1x}=0 \qquad (26)$$

so that in this case only one component[4] of vector potential A_ϕ is different from zero, the others, A_r and A_x, being equal to zero.

When written in this form, the results can be visualized as in Figure 3. Here for a point charge at P the scalar equipotential surfaces are spheres given by $V=1/R$. The vector equipotential lines are given by the intersection of these spheres with any cone whose vertex is at P. This follows from equation 26 since on the spherical surface R is a constant, on the line of intersection between the cone and sphere $r^2=y^2+z^2$ is a constant, and x is the perpendicular distance from P to the circular disk whose boundary is the circle A_ϕ=constant. It is also evident from the symmetry of the figure. Many authors simplify the case of circular symmetry still further by using as vector potential $A_{1\phi}=rA_\phi$ instead of A_ϕ given in equation 26. This involves a slight change in the definition of vector potential but gives the same value of flux density B by taking the curl in a slightly different way. It has the advantage of making the cone with vertex at P a vector equipotential surface for $A_{1\phi}$. This can be seen from the fact that the numerator and denominator of $A_{1\phi}$ are of the same dimensions. Thus, for any given cone, as R is increased, x, y, and z are increased proportionally and the value of $A_{1\phi}$ thus remains constant.

From this it can be shown further that the vector potential $A_{1\phi}$ at any point on the circle in this case is proportional to the solid angle which the circle subtends at the origin. This latter property also is used by Lamb[4] on page 127. It also shows that the vector potential along the x axis, taken as the vertical, is zero. Since, however, this is an arbitrary choice and another zero for the vector potential might have been chosen, this illustrates why the definition of vector potential is not unique.

This result was derived without the use of the tables since the potential $V=1/R$ for a point source was known. As a first example in the use of the tables, let us find expressions for the same quantities. To use the tables the differential equation to be solved is first set down. It is

$$\nabla^2V=\rho \qquad (27)$$

with the boundary conditions

$$V=0 \text{ when } x=-\infty,\ V=0 \text{ when } y=-\infty,\ V=0 \text{ when } z=-\infty \qquad (28)$$
$$V=0 \text{ when } x=\infty,\ V=0 \text{ when } y=\infty,\ V=0 \text{ when } z=\infty$$

and ρ corresponds to a unit point charge at the point x_1, y_1, z_1. The solution of equations 27 and 28 is found under Section III of the tables as follows[5]

$$V=X_{1p}(-\infty,\infty)Y_{1q}(-\infty,\infty)\times Z_{1u}(-\infty,\infty)/(\alpha^2+\beta^2+\gamma^2) \qquad (29)$$

where

$$X_{1p}(-\infty,\infty)=\frac{1}{\pi}\int_0^\infty \cos\alpha(x-x_1)d\alpha$$
$$Y_{1q}(-\infty,\infty)=\frac{1}{\pi}\int_0^\infty \cos\beta(y-y_1)d\beta \qquad (30)$$
$$Z_{1u}(-\infty,\infty)=\frac{1}{\pi}\int_0^\infty \cos\gamma(z-z_1)d\gamma$$

Using equation 3 we find

$$A_{2x1}=-\int_\infty^y dy\frac{\partial}{\partial z}X_{1p}(-\infty,\infty)\times Y_{1q}(-\infty,\infty)Z_{1u}(-\infty,\infty)/(\alpha^2+\beta^2+\gamma^2)$$
$$=\frac{1}{\pi^3}\int_0^\infty d\alpha\int_0^\infty d\beta\int_0^\infty d\gamma\left(\frac{\gamma}{\beta}\right)\times \cos\alpha(x-x_1)\sin\beta(y-y_1)\times \sin\gamma(z-z_1)/(\alpha^2+\beta^2+\gamma^2)$$
$$A_{2y1}=0 \qquad (31)$$
$$A_{2z1}=\int_\infty^y dy\frac{\partial}{\partial x}X_{1p}(-\infty,\infty)Y_{1q}\times (-\infty,\infty)Z_{1u}(-\infty,\infty)/(\alpha^2+\beta^2+\gamma^2)$$
$$=\frac{1}{\pi^3}\int_0^\infty d\alpha\int_0^\infty d\beta\int_0^\infty d\gamma\left(\frac{\alpha}{\beta}\right)\times \sin\alpha(x-x_1)\sin\beta(y-y_1)\times \cos\gamma(z-z_1)/(\alpha^2+\beta^2+\gamma^2)$$

These expressions for A_{2x1}, A_{2y1}, and A_{2z1} are the same as those given for A_{2x}, A_{2y}, and A_z, in equation 23 with the exception that they may differ by a constant, since in the integration here with respect to y we omitted the term corresponding to the lower limit ∞ and this would only add a constant to the solution.

It would appear at first sight that the expressions given in equation 23 are much simpler to use than those given in equation 31. This is true in this particular case of a point charge. However, when it is necessary to consider a complete circuit so that integration of equation 23 or 31 is required, it generally is easier to use equation 31 as shall be shown now.

The Scalar and Vector Potential Due to Thin Rectangular Sheet of Magnetic Material or a Rectangular Coil Using Definition III

A thin sheet of magnetic material in the yz plane is bounded by the lines $y=a$, $y=-a$, $z=b$, $z=-b$. It is magnetized uniformly in the x direction. Find the scalar and vector potentials.

The plate may be regarded as made up of infinitesimal magnets of area $dy\,dz$ each of which consists of a doublet in the x direction. The field V_D due to a doublet in the x direction can be derived from the field due to a point charge by differentiating it with respect to x since by definition

$$V_D = \operatorname*{Limit}_{\Delta x_1 \to 0} \times \frac{V(x,y,z;\ x_1,y_1,z_1) - V(x,y,z;\ x_1+\Delta x_1,y_1,z_1)}{\Delta x_1} = \frac{\partial V}{\partial x} \tag{32}$$

Then integrating $\partial V/\partial x$ from $y=-a$ to $y=+a$ and from $z=-b$ to $z=b$, we obtain for the required potential

$$V_1 = -\frac{1}{\pi^3}\int_{-a}^{a} dy_1 \int_{-b}^{b} dz_1 \int_0^\infty d\alpha \times \int_0^\infty d\beta \int_0^\infty \alpha d\gamma \sin \alpha(x-x_1) \times \cos \beta(y-y_1) \cos \gamma(z-z_1)/\times (\alpha^2+\beta^2+\gamma^2) \tag{33}$$

and from this using equation 10

$$A_x = -\frac{1}{\pi^3}\int_{-a}^{a} dy_1 \int_{-b}^{b} dz_1 \int_0^\infty d\alpha \times \int_0^\infty d\beta \int_0^\infty d\gamma\left(\frac{\alpha\gamma}{\beta}\right) \sin \alpha(x-x_1) \times \sin \beta(y-y_1) \sin \gamma(z-z_1)/\times (\alpha^2+\beta^2+\gamma^2) \tag{34}$$

$$A_y = 0 \tag{35}$$

$$A_z = \frac{1}{\pi^3}\int_{-a}^{a} dy_1 \int_{-b}^{b} dz_1 \int_0^\infty d\alpha \times \int_0^\infty d\beta \int_0^\infty d\gamma\left(\frac{\alpha^2}{\beta}\right) \cos \alpha(x-x_1) \times \sin \beta(y-y_1) \cos \gamma(z-z_1)/\times (\alpha^2+\beta^2+\gamma^2) \tag{36}$$

These are the required expressions for scalar and vector potential. They can be simplified by performing some of the integrations, but it will not be done here.

To find the field due to a rectangular coil with the boundaries outlined, it will be recollected there is a well-known theorem.[6] "A current flowing in any closed circuit produces the same magnetic field as a certain magnetic shell known as the equivalent magnetic shell. This shell may be taken to be any shell having the circuit for its boundary, its strength being uniform and proportional to that of the current." From this it can be concluded that the magnetic field produced by the rectangular-shaped thin magnetic sheet is similar to that produced by the current in a wire which coincides with the boundary of the thin magnetic sheet. Thus the scalar and vector potentials in equations 33 to 36 apply to the field of the rectangular coil.

The Scalar and Vector Potential Due to a Rectangular Coil Using Definition II

The magnetic vector potential for a rectangular coil can be obtained directly from Definition II without using the preceding method of replacing it by an equivalent magnetic shell. Since this procedure is sometimes more useful it will be illustrated for the same coil as in the previous example.

Consider first the two sides of the coil at $y=-a$ and $y=a$. The current may be considered to be made up of a number of moving charges at each element dz of the coil so that the equations for scalar and vector potential are as shown in equations 37

$$\begin{aligned} \nabla^2 V &= -\rho \\ \nabla^2 A_x &= 0 \\ \nabla^2 A_y &= \rho u_y/c \\ \nabla^2 A_z &= \rho u_z/c \end{aligned} \tag{37}$$

with the boundary conditions for V

$$V=0 \text{ when } x=-\infty,\ V=0 \text{ when } y=-\infty,\ V=0 \text{ when } z=-\infty \tag{38}$$

$$V=0 \text{ when } x=+\infty,\ V=0 \text{ when } y=+\infty,\ V=0 \text{ when } z=+\infty$$

with similar boundary conditions for A_x, A_y, and A_z. If ρ corresponds to a point charge, then the solution of these equations is found from Section III of the tables to be

$$V = X_{1p}(-\infty,\infty)Y_{1q}(-\infty,\infty)\times Z_{1u}(-\infty,\infty)/(\alpha^2+\beta^2+\gamma^2) \tag{39}$$

where $X_{1p}(-\infty,\infty)\,Y_{1q}(-\infty,\infty)\ \ Z_{1u}(-\infty,\infty)$ are given in equation 30.

Since the coil is in the yz plane there is no component of current in the x direction and thus

$$A_x=0,\ A_y=u_yV/c,\ A_z=u_zV/c \tag{40}$$

Hence the vector potential due to such elements extending from $z=-b$ to $z=b$ is found by integrating with respect to z from $-b$ to b giving

$$A_z = \frac{u_z}{\pi^3}\int_{-b}^{b} dz_1 \int_0^\infty d\alpha \int_0^\infty d\beta \int_0^\infty d\gamma \times \cos \alpha(x-x_1) \cos \beta(y-y_1) \times \cos \gamma(z-z_1)/c(\alpha^2+\beta^2+\gamma^2) \tag{41}$$

However, there are two such sides of the coil with current in opposite directions, one at $y_1=-a$ and the other at $y_1=a$ and thus the vector potential due to this is given by

$$A_z(-a,a) = -\frac{u_z}{\pi^3}\int_{-b}^{b} dz_1 \int_0^\infty d\alpha \int_0^\infty d\beta \times \int_0^\infty d\gamma \frac{\cos \alpha(x-x_1)[\cos \beta(y+a) - \cos \beta(y-a)] \cos \gamma(z-z_1)}{c(\alpha^2+\beta^2+\gamma^2)} \tag{42}$$

Similarly, the vector potential due to the two coil sides at $z=-b$ and $z=b$ is given by

$$A_y(-b,b) = \frac{u_y}{\pi^3}\int_{-a}^{a} dy_1 \int_0^\infty d\alpha \int_0^\infty d\beta \times \int_0^\infty d\gamma \frac{\cos \alpha(x-x_1) \cos \beta(y-y_1) \times [\cos \gamma(z+b) - \cos \gamma(z-b)]}{c(\alpha^2+\beta^2+\gamma^2)} \tag{43}$$

Thus the vector potential due to the coil is given by $A_y(-b,b)$, $A_z(-a,a)$ together with $A_x=0$. It can be shown that these give the same values for B_x, B_y, and B_z as the vector potential in equations 34, 35, and 36 on recognizing that

$$\int_0^\infty d\alpha \int_0^\infty d\beta \int_0^\infty d\gamma \cos \alpha(x-x_1) \times \cos \beta(y-y_1) \cos \gamma(z-z_1) = 0 \tag{43A}$$

except at $x=x_1$, $y=y_1$, and $z=z_1$.

Vector Potential of a Circular Coil

The vector potential due to a circular coil of radius a can be calculated in a similar manner. Since the published tables of Green's Functions cover only rectangular co-ordinates, it is necessary to use other known results. We have in cylindrical co-ordinates.[7]

$$\frac{1}{R} = 2\int_0^\infty \epsilon^{-k(z-z_1)} \sum_m{}' J_m(kr)J_m(kr_1) \times \cos m(\phi-\phi_1)dk \tag{44}$$

where $R^2=r^2+r_1^2-2rr_1 \cos(\phi-\phi_1)+(z-z_1)^2$. Thus the expression gives the scalar potential at r, z, ϕ due to a point charge at r_1,ϕ_1,z_1.

Consider the field due to the circular coil as replaced by the equivalent magnetic shell of a thin flat disk of magnetic material bounded by the circle of radius a. The potential due to such a disk can be obtained as before by considering it to be composed of small doublet elements of area $r_1d\phi_1dr_1$. The field due to a doublet in the z direction is found from equation 44 by differentiating it with respect to z. Thus the potential due to the magnetic disk is given by

$$\begin{aligned} V_{ma} &= \int_0^a dr_1 \int_0^{2\pi} r_1d\phi_1 \frac{d}{dz_1}\left(\frac{1}{R}\right) \\ &= 2m\int_0^a dr_1 \int_0^{2\pi} r_1d\phi_1 \frac{d}{dz_1} \times \int_0^\infty \epsilon^{-k(z-z_1)} \sum_m{}' J_m(kr) \times J_m(kr_1) \cos m(\phi-\phi_1)dk \\ &= 2m\pi a \int_0^\infty \epsilon^{-k(z-z_1)} J_0(kr)J_1(ka)dk \end{aligned} \tag{45}$$

This gives the scalar potential. The vector potential is found by a procedure similar to equation 12 but now in cylindrical co-ordinates it is found to be

$$A_\phi=2m\pi ar\int_0^\infty \epsilon^{-k(z-z_1)}J_1(kr)J_1(ka)dk$$

$$A_r=0$$

$$A_z=0$$

Both of these expressions may be checked with known results[8] given by Lamb. While Lamb's work applies to hydrodynamics, it corresponds as far as the mathematical procedure goes to the present problem when Lamb's velocity potential corresponds to our scalar potential and his stream function corresponds to our vector potential. He gives a drawing showing the stream lines in the hydrodynamics case which corresponds to our vector potential. Maxwell[9] also gives an illustration showing the lines of force and scalar equipotential surfaces for the case of a circular coil. His lines of force are the same as Lamb's stream lines.

The Effect of Variation of the Field with Time

It may be well to point out that although the definitions of vector potential given apply to cases where the fields change with time, the examples have been limited to cases where the fields do not change with time. The tables of Green's Functions can be applied equally well to those cases where the fields vary with time.

However, the equation $E_z=\partial V/\partial_z$ then must be replaced by the more complex one of $E_z=\partial V/\partial x+\mu\partial A_z/\partial t$. Also, since there is variation with time, the velocity of propagation appears giving the results in the form of waves which are quite similar to wave guide theory. This requires the introduction of some additional concepts which are too lengthy to cover in the present paper. Therefore, it seemed best to treat the simpler case of no variation with time first, since this covers a large percentage of the problems usually encountered and is a type of solution which is given in existing literature. Having thus shown how the procedure works in these cases, the next step would be to consider the more general case of variation with time when there is little in the way of existing results in the literature which may be used as a guide. It will be found that by following the procedure outlined here for the case of no variation with time, the consideration of cases where there is variation with time is obtained without difficulty from a further development of this procedure. It will be covered in a later paper.

Conclusion

It has been shown how the tables of Green's Functions previously published give the solution in problems where it is desired to use vector potential. The various definitions of vector potential have been compared. It is known that the definition of vector potential is not unique so that two different vector potentials may give rise to the same value of flux density B and it was shown how one such vector potential may be converted into the other. Several examples have been worked out obtaining both the scalar and vector potential in the case of a point source, a rectangular and a circular coil, and a thin rectangular and a thin circular disk of magnetic material.

Appendix I

The field due to an infinite wire provides an example of scalar and vector potentials in terms of simple expressions which may be helpful in understanding the relations between these quantities in the 2-dimensional case. For brevity only the results are given, the details being omitted. Two different cases will be considered.

Case I

For the first case assume the infinite wire has a static charge on it.

The electrical field can be derived from the scalar potential

$$V=2\log r/r_0$$

or from the vector potential

$A_r=0$, $A_\theta=0$, and $A_z=2\theta$

Case II

For the second case assume the infinite wire has a current in it.

The magnetic field can be derived from the scalar potential

$$V=2\theta$$

or from the vector potential

$A_r=0$, $A_\theta=0$, and $A_z=2\log r/r_0$

In both cases the differentiation with respect to r is of the form $\partial/\partial r$ and that with respect to θ is of the form $\partial/r\partial\theta$. By comparing these two cases it is seen that for a static charge on the infinite wire the scalar equipotential lines are given by $r=$constant and the vector equipotential lines are given by $\theta=$constant which also gives the lines of force.

When the field is caused by current in the wire the scalar equipotential lines are given by $\theta=$constant and the vector equipotential lines by $r=$constant which again gives the lines of force. It is obvious that in this 2-dimensional example the field configuration is the same in both cases. However, the curves of V and A are interchanged.

References

1. TABLES OF GREEN'S FUNCTIONS, FOURIER SERIES, AND IMPULSE FUNCTIONS FOR RECTANGULAR CO-ORDINATE SYSTEMS, J. J. Smith. *AIEE Transactions*, volume 70, part I, 1951, pages 22–30.

2. AMERICAN STANDARD DEFINITIONS OF ELECTRICAL TERMS. *ASA C42–1941*, *AIEE*, 1941, section 05.05.150.

3. ELECTROMAGNETIC RADIATION (book), G. A. Schott. Cambridge University Press, Cambridge, England, 1912, pages 3–4.

4. HYDRODYNAMICS (book), H. Lamb. Cambridge University Press, Cambridge, England, 6th edition, 1932, page 138. In equation 9 the same expression is given.

5. It can be shown independently that this expression for $V=1/R$. See OPERATIONAL CALCULUS (book), B. Van Der Pol, H. Bremmer. Cambridge University Press, Cambridge, England, 1950, page 359.

6. ELECTRICITY AND MAGNETISM (book), J. H. Jeans. Cambridge University Press, Cambridge, England, 2nd edition, 1911, page 426.

7. BESSEL FUNCTIONS (book), A. Gray, G. B. Mathews, T. M. MacRobert. The MacMillan Company, New York, N. Y., 2nd edition, 1922, page 102, case I where $V=1/R$.

8. HYDRODYNAMICS (book), H. Lamb. Cambridge University Press, Cambridge, England, 6th edition, 1945, page 239, paragraph 161.

9. ELECTRICITY AND MAGNETISM (book), J. C. Maxwell. Oxford University Press, London, England, 3rd edition, volume II, 1904, figure XVIII.

No Discussion

Typical Block Diagrams for a Transistor Digital Computer

J. H. FELKER
NONMEMBER AIEE

Synopsis: The first electric digital computers were built around the properties of relays. The superior speed capabilities of vacuum tubes has led in recent years to their use in new computer designs to replace relays. Because of the small size, low power consumption, and expected long life of transistors, it now appears that the transistor will replace the vacuum tube as a computer element. This paper presents a study of binary computer functions with recommended mechanizations that were selected because they appeared to be readily attainable with transistors now under development. Block diagrams are presented of switches, memory units, arithmetic units, and other basic components. Estimates are given for the number of parts required in the units. It is concluded that a high-performance all-semiconductor computer can be built with germanium diodes and transistors.

DIGITAL computer operations[1] can be divided into two classes, memory and logic. Memory includes the three operations write, read, and erase, where the write operation, for example, can be defined as the representation in space of a function of time. In computers, the logic operations consist of the recognition of spatial distributions of voltages and currents. All the logic operations that can be described in words can be mechanized with suitable combinations of three basic functions. The basic functions are "or," "and," and "inhibition" or negation. An *n*-terminal or-circuit will have an output when any one of its *n* input leads is energized. An *n*-terminal and-circuit will have an output only if all of its *n* leads are energized. Inhibition leads can be added to either of the two circuits with the result that the circuit cannot have an output as long as there is a signal on any of the inhibition leads. These logic circuits have the advantage that they can be built out of passive nonlinearities. A nonlinear device has two states, a high-impedance and a low-impedance state. The more pronounced the nonlinearity the more satisfactory the unit is as a 2-state element. Since nonlinearity can be obtained from purely passive elements, it is possible to obtain all of the logic functions of a computer without the use of vacuum tubes or other active elements. In practice, it is necessary to use some active elements because the nonlinear devices available are not ideal and have losses. It is necessary, therefore, to use amplifiers to make up for losses in the logic circuits. It is important to note that as the efficiency of the logic circuits is increased, the number of amplifiers required is decreased.

This computer design philosophy was followed in the design of the National Bureau of Standards Computer SEAC.[2] It is believed that the approach that will result in a vacuum-tubeless computer at the earliest date is to follow the SEAC example, in so far as the use of germanium diode logic circuits is concerned, but replacing the vacuum-tube amplifiers with transistor amplifiers. Since the transistor itself has voltage and current relationships quite similar to a germanium diode, it is expected that the diodes in a transistor computer will operate in a more natural environment than in a vacuum-tube computer and will respond to such favorable conditions by exhibiting longer life and more reliable operation. It is for such an all-semiconductor computer that the block diagrams described herein have been planned.

Serial operation has been used in all of these designs. In serial operation all the digits of a word or number are transmitted in series along a single wire with the least significant digit first. The interval between the rise of successive digit pulses is referred to as one digit time. The interval between successive numbers or words is commonly referred to as the word time. Digit times of 1 microsecond have proved to be practicable, and words as long as 50 binary digits have been employed (equivalent to 15 decimal digits).

A number of the block diagrams presented herein have been mechanized with a transistor amplifier that regenerates pulses at a megacycle rate. This amplifier operates on a total power drain of approximately 50 milliwatts. The details of this amplifier and the associated logic circuits have been presented by the writer in another paper.[3]

The reasons for choosing serial instead of parallel operation are quite simple. In a parallel system, each digit of a number appears in a different circuit. In a parallel adder a separate adder stage is used for each digit, whereas in a serial adder only one adder stage is used, and this one stage handles all the digits. Not only are the arithmetic operations performed with fewer components in the serial machine but switching is also simplified. To switch an *n*-digit parallel number requires *n* switches to switch the separate wires, whereas in the serial machine only one switch is required.

Basic Building Blocks

The proposed computer elements are based on an assembly of a relatively small number of different kinds of basic units. The basic units have been limited to an or-circuit, an and-circuit, an inhibitor-circuit, an amplifier, and a storage cell based on a delay line. The designs suggested are similar to those in the EDVAC designed at the University of Pennsylvania and the Bureau of Standards SEAC. Some of the basic components have been described in in the literature.[4,5] They are described again in the following paragraphs to provide a specific basis for the estimates of numbers of parts that are a goal of this study.

An *n*-terminal or-circuit (see Figure 1) develops an output when any one of its input terminals is energized. Crystal diodes are put in series with each input to prevent a pulse at one input from feeding back to any of the other inputs. An *n*-terminal circuit, therefore, requires *n* crystal diodes.

An and-circuit having *n* input ter-

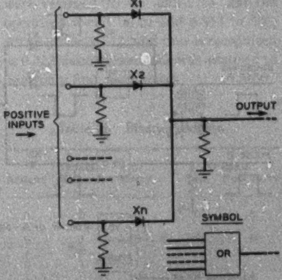


Figure 1. Or-circuit

Paper 52-162, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted October 31, 1951; made available for printing April 17, 1952.

J. H. FELKER is with Bell Telephone Laboratories, Whippany, N. J.

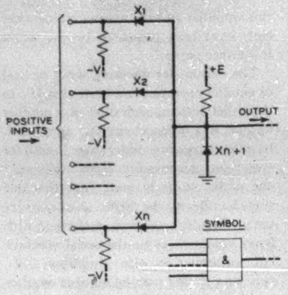


Figure 2. And-circuit

minals develops an output only when all n of the input terminals are energized. In Figure 2 each of the inputs is returned to a negative voltage and the output is clamped slightly below ground by X_{n+1}. Only when all of the inputs rise above ground will X_{n+1} be cut off, permitting the output to rise. Thus the output consists of the overlapping part of the inputs. An n-terminal circuit is seen to require $n+1$ crystals.

Two terminal and-circuits are used extensively in serial computers for retiming signals. One terminal of the circuit is fed by the signal to be retimed while the other is fed by digit pulses from a master clock. There will be an output from the circuit only for the overlap of the master digit pulse and input signal. The circuit whereby the output is made a replica of the pulse from the clock is discussed in connection with the basic amplifier.

An inhibitor terminal can be added to any and-circuit or or-circuit. Such a circuit operates as though there were no inhibitor terminal when the inhibiting pulse is absent. When the inhibiting pulse is present, however, the circuit prevents any output from being developed. The inhibiting circuits used in the designs proposed herein are of the simple variety shown in Figure 3 where positive inputs synchronized in time are required. Note that the signal to be inhibited is passed through an eighth-digit delay line, while the inhibiting pulse is passed both through and around a quarter-digit delay line. This insures that the inhibitor pulse will, in effect, arrive earlier than the signal pulse and last longer. In the absence of input pulses, crystal diode X_4 will clamp the output at ground because input B is returned to a negative potential. Note that X_1 and X_2 are returned through the transformer to a positive potential. If input B goes positive (without an inhibiting pulse appearing on input A) X_4 will be cut off and the output voltage will rise until it is clamped at the positive potential to which X_1 and X_2 are returned. If there is an inhibiting pulse (positive) it is inverted by the transformer and will carry X_1 and X_2 negative, which will keep X_4 conducting, no matter what happens at B. Thus, if pulses A and B were written as a 2-digit binary number AB, the circuit translates 01 into a 1 at the output. It translates 00, 10, and 11 into zero at the output.

Active elements will be used not as flip-flops or switches but as repeating amplifiers to make up for attenuation in crystal diode circuits and delay lines. The standard use will include a retiming feature as well as amplification. Wherever a pulse is likely to suffer intolerable attenuation, deformation, or a variable delay, a circuit like that of Figure 4 is inserted in the machine.

The assembly shown has two inputs, A and B. Input A is the pulse to be retimed and amplified. Input B comes from the master clock. This component supplies reference pulses (known as digit pulses) every digit time. These pulses are available in various phases, that is, with various but accurately controlled delays of a fraction of a digit time. The pulse fed to B is selected to rise sometime between the expected rises and falls of the pulses on A.

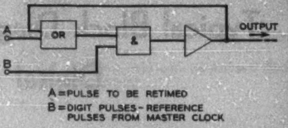


Figure 4. Amplifier with pulse retiming

If there is no input on A, there will be no output from the amplifier because of the and-circuit. If there is an input on A, when the digit pulse arrives the amplifier output will rise with a rise time determined by the digit pulse (assuming that the amplifier passband does not limit it). Part of the amplifier output is fed back through an or-circuit to the and-circuit. This insures that the output pulse will not fall until the reference digit pulse does, even though pulse A may have ended after B rose.

This reshaping with logic circuits and an amplifier is the way in which every pulse is maintained with the desired time synchronization. Pulse A may vary somewhat in the delay it has suffered but the output pulse still will leave the amplifier at a time determined only by the reference pulse from the master clock. Thus, the pulses in the computer are made to have fixed durations and to occur at designated times.

In the tabulations of components in the following paragraphs an active element is intended to represent as one unit a vacuum tube with transformer or a transistor. The term will include neither the and-circuit nor the or-circuit shown in Figure 4.

The basic storage cell proposed is not a static device like a flip-flop but is an electric delay line plus an amplifier. When vacuum tubes are used, this type of storage saves one active element in a 1-digit storage cell and is believed to be a more reliable use of active elements. In larger storage units more elements will be saved.

A block diagram of a cell is shown in Figure 5. The unit has three inputs: digit pulses, the signal to be stored, and an erase pulse. The digit pulses are received from the master clock every digit time and are used to retime the output of the

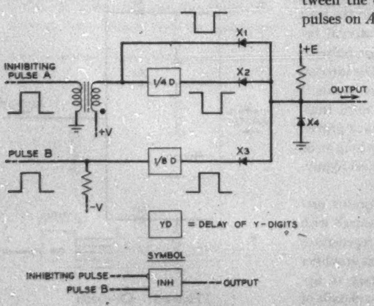


Figure 3 (left). Inhibitor-circuit

Figure 5 (right). Storage cell

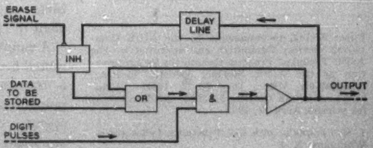


Table I. Basic Building Blocks

Unit	Crystal Diodes	Digits of Delay	Active Elements
N-terminal or -circuit	N	0	0
N-terminal and -circuit	N+1	0	0
Inhibitor-circuit	4	$^3/_8$	0
N-digit storage cell	10*	N	1*

* One active element will be used for $N \leq 8$; for $N \geq 8$ the number of active elements planned is the smallest integer $\geq N/8$. Six additional crystal diodes will be required with each active element.

delay line before it is amplified and recirculated. The erase signal is received whenever new data are to be stored. It serves to erase the data in storage, blocking the delay line output from its input until the new data have been inserted.

The delay line may be long enough to store one word or just one digit of data. It is believed that up to 15-digit delay lines with lumped impedances can be built that will hold the delay constant to within a small fraction of one digit time. Depending on the length of a word, it may be necessary to break 1-word lines into sections and insert an amplifier between sections to retime and regenerate the pulses stored. To insure conservative estimates, the estimates made in following sections are on the basis of regeneration after every eight digits of delay.

The circuits which have been discussed use the components listed in Table I. This table will be used frequently in estimating the parts required in the larger assemblies.

Switches

The switches are planned to combine a switching and a storage function. When a switch is given instructions to go to the k'th position, it goes there and it remembers that it is to remain there (self-locking operation). All the elements of the switches have been discussed in preceding paragraphs. How these elements are combined to make switches is described in the following:

A single-pole double-throw switch, as shown in Figure 6, consists of a storage cell and a switch unit. When a 1 is stored in the storage unit, as the result of a pulse on the switching instruction lead, the left-hand and-circuit of the switch unit will pass signal a while the inhibitor blocks signal b. When a zero is stored, the and-circuit will block signal a and the inhibitor-circuit in the switch unit will pass signal b.

Whenever the switch is to be reset, an erase signal is fed to the storage unit, which then drops its old instruction and goes to position b unless the new instruction sets it to a.

The switch unit itself uses 13 crystal diodes, 3/8 digit of delay, and one active element, as can be ascertained from Figure 6 and Table I. A complete single-pole double-throw switch has 23 diodes, $1^3/_4$ digits of delay, and two active elements. A 2-pole double-throw switch would have two switch units and one storage unit. A 3- or 4-pole switch would have three or four switch units, one storage unit, and perhaps an extra amplifier to prevent the switch units from loading down the storage unit excessively.

Switches with more than two positions can be assembled in a slightly different manner from the double-throw switches. Suppose, for discussion, that an 8-position switch is needed. A 3-digit code must be sent to the switch to specify the position it is to select. A convenient way to operate such a switch is to translate the 3-digit code into a 6-wire code so that each digit is represented by signals of opposite phase on a pair of wires. In this system, a 1 is represented by a positive pulse on the positive bus and a negative pulse on the negative bus. A zero is represented by a negative pulse on the positive bus and a positive pulse on the negative bus. A 3-terminal and-circuit is provided for each of the eight lines that may be selected. The and-circuits are operated by positive pulses, and all are connected to either one or the other of each pair of wires representing the three digits of the position code. The and-circuit on the fifth (101) wire, for example, is connected to the positive bus of the pair representing the coefficient of 2^2, to the negative bus of 2^1, and the positive bus of 2^0. The three leads of the and-circuit will be energized by positive pulses only when 101 is sent to the switch as instructions.

A single-pole 8-position switch is shown in Figure 7. The eight 3-terminal and-circuits (one for each horizontal lead), plus the 9-terminal or-circuit make an easily studied crystal matrix which feeds another and-circuit for retiming of pulses and an amplifier on the output. The and-circuits are fed by switches S_0, S_1, and S_2, which are driven by the switching instructions. These three switches are similar to the one shown in Figure 6. S_0 is operated by the coefficient of 2^0, S_1 by the coefficient of 2^1, and S_2 by the coefficient of 2^2. As the figure is drawn, only the and-circuit on line 7 (111) is energized. A pulse will be received at the output whenever a pulse is put on line 7. Pulses can be put on any of the other lines without getting to the output, because every other line is held negative by at least one lead from S_0, S_1, or S_2. If an erase signal were sent to the three storage cells of S_0, S_1, and S_2, the number 111 would be erased, leaving 000, and the switches would be free to move to new positions and select another one of the eight inputs.

Table II. Switches

Type of Switch	Crystal Diodes	Digits of Delay	Active Elements
1P.2T	23	$1^3/_4$	2
2P.2T	36	$1^3/_4$	3
4P.2T	63	$6^1/_2$	6
1P.4T	61	$2^3/_4$	4
1P.8T	105	6	7
1P.16T	160	7	9
1P.32T	311	$8^3/_4$	11

Where *2P.2T*, for example, represents a 2-pole 2-throw switch.

Based on this type of mechanization, an r-position switch requires n single-pole double-throw switches, where n is the smallest integer equal to or greater than $\log_2 r$. The switch requires $r\ (n+1)$ crystal diodes, plus four diodes for retiming the output, and an output amplifier.

Table II gives estimates of the crystal diodes, digits of delay, and active elements required in switches of different degrees of complexity.

Handling of Negative Numbers

The most significant digit place of every number is reserved to indicate the sign of the number. Positive numbers have a zero in the last place. A negative number is obtained by taking the 2's complement of the positive number. This results in every negative number having a 1 in its last place. The system is equivalent to the 10's complement method used in decimal calculators. In a decimal calculator operating with three significant figures, a fourth place might be provided for the sign. The number −187 might be represented by its 10's complement 9,813. Then, for example, if −187 were required to be added to 500 the operation would be to add 9,813 to 500, which gives 10,313, and thus is recognized as 0,313 since the machine was assumed to have only four digit places.

Table III. Binary Addition

	Inputs		Outputs	
Addend	Augend	Carry	Sum	New Carry
0	0	0	0	0
0	0	1	1	0
0	1	0		
1	0	0		
0	1	1	0	1
1	1	0		
1	0	1		
1	1	1	1	1

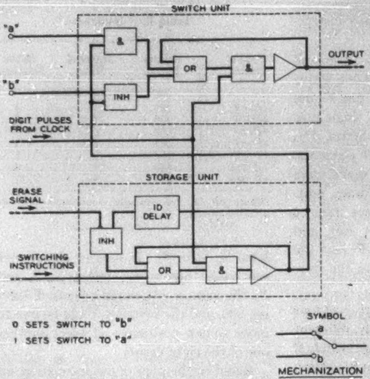
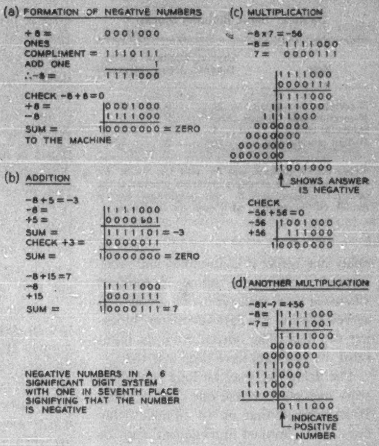

SWITCH UNIT
"a"
"b"
&
INH
OR
&
OUTPUT
DIGIT PULSES FROM CLOCK
STORAGE UNIT
ERASE SIGNAL
ID DELAY
INH
SWITCHING INSTRUCTIONS
OR
&
0 SETS SWITCH TO "b"
1 SETS SWITCH TO "a"
SYMBOL
a
b
MECHANIZATION

Figure 6 (left). A simple switch

Figure 8 (right). Handling of negative numbers


(a) FORMATION OF NEGATIVE NUMBERS
+8 = 0001000
ONES COMPLIMENT = 1110111
ADD ONE 1
∴ −8 = 1111000
CHECK −8+8=0
+8 = 0001000
−8 1111000
SUM = 1 0000000 = ZERO
TO THE MACHINE
(b) ADDITION
−8+5=−3
−8 = 1111000
+5 = 0000101
SUM = 1111101 = −3
CHECK +3 = 0000011
SUM = 1 0000000 = ZERO
−8+15=7
−8 1111000
+15 0001111
SUM = 1 0000111 = 7
NEGATIVE NUMBERS IN A 6 SIGNIFICANT DIGIT SYSTEM WITH ONE IN SEVENTH PLACE SIGNIFYING THAT THE NUMBER IS NEGATIVE
(c) MULTIPLICATION
−8×7=−56
−8= 1111000
7= 0000111
1111000
0000111
1111000
1 111000
11 11000
000 0000
0000 000
00000 00
000000 0
1001000
SHOWS ANSWER IS NEGATIVE
CHECK
−56+56=0
−56 1001000
+56 111000
1 0000000
(d) ANOTHER MULTIPLICATION
−8×−7=+56
−8= 1111000
−7= 1111001
1111000
0 000000
00 00000
111 1000
1111 000
1110 00
11100 0
0111000
INDICATES POSITIVE NUMBER

A negative number (2's complement) can be obtained in the binary computer by first, forming the 1's complement (changing all zeroes to 1's and vice versa by means of an inhibitor circuit), and then adding 1. Figure 8 shows several examples of binary arithmetic performed with negative numbers.

A point of interest is that even though either or both the multiplicand and multiplier are negative the correct sign will be obtained for a product provided the negative numbers are increased from their normal length W to $2W$-1 by filling in 1's before multiplying. This is necessary because the product of two W-digit numbers, where the last digit specifies sign, is a number $2W$-1 digits long. Unless the multiplicand and multiplier are increased to length $2W$-1 (when negative) the $2W$-1 place may be incorrect. This lengthening is not required for positive numbers, because the digits from W to $2W$-1 would be zeroes if they were filled in and would contribute nothing to the product. A consequence is that the product XY can be obtained to $2W$-1 places with $W(W$-1) elementary additions provided X and Y are positive while $W(2W$-1) additions will be required if both are negative. Where fast multiplication is desired it may be advisable to convert negative numbers to positive ones, multiply, and then adjust the sign of the product.

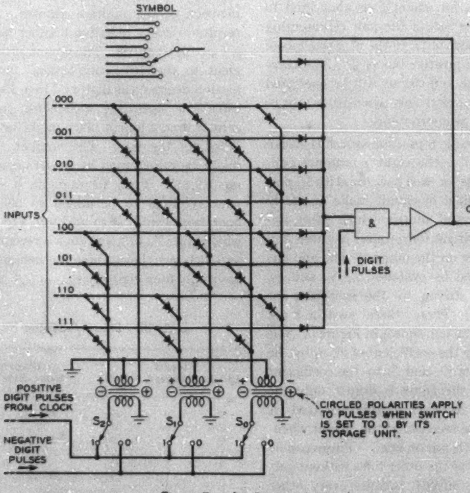

SYMBOL
000
001
010
011
100
101
110
111
INPUTS
&
OUTPUT
DIGIT PULSES
POSITIVE DIGIT PULSES FROM CLOCK
NEGATIVE DIGIT PULSES
+
−
S2
S1
S0
1
0
CIRCLED POLARITIES APPLY TO PULSES WHEN SWITCH IS SET TO 0 BY ITS STORAGE UNIT.

Figure 7. An 8-position switch

Adder

The adder can be considered as a translator with three inputs: addend, augend, and carry. It is a simple translator in that its output is a function only of the number of ones among its three inputs, as can be seen from Table III.

The combination 0 0 0 is automatically taken care of in the adder shown in Figure 9. The three dashed circuits at the left of the block diagram recognize three situations among the three inputs. The situations are: at least one 1, at least two 1's and three 1's among the inputs. If there is only one 1, it will go through the bottom or-circuit, the following inhibitor circuit, and then another or-circuit. After being reclocked and amplified it will provide a 1 as the sum. In this case none of the and-circuits on the A, B, and carry leads will have operated. If there are at least two 1's on the A, B, and carry leads, at least one of the three 2-terminal and-circuits in the dashed box will operate, with two results. The output of the 3-terminal or-circuit at the bottom left of the diagram will be inhibited so that it makes no contribution to the sum. In addition, a carry signal will be developed which is delayed one digit, reclocked, and amplified to serve as the carry for the

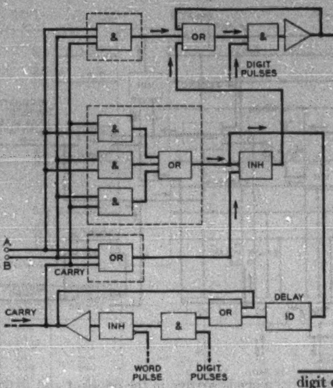


Figure 9 (left). Adder

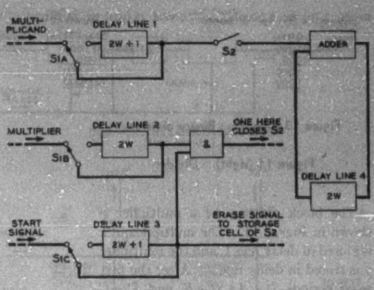


Figure 11 (right). Multiplier

next augend and addend. If there are three 1's, the 3-terminal and-circuit at the top and left-hand side of the diagram will operate and develop a sum of one. The three 2-terminal and-circuits on the left of the diagram also will have operated and provided the carry. Thus, the adder table is mechanized with two active elements (amplifiers).

The inhibitor circuit in series with the carry lead should be noted. This circuit is fed by a word pulse as well as the carry digits. The word pulse is received in synchronism with the first digit of every number. The word pulse will inhibit the carry pulse if one is present and will prevent a carry developed in one problem from being used in the next. This feature is required in the addition of negative numbers.

The carry lead is brought outside the adder to facilitate subtraction. Suppose x is to be subtracted from y. The number y might be fed to the addend terminal and the number x fed through an inhibitor circuit to the augend terminal. The inhibitor would also be fed by digit pulses from the master clock, and the augend would therefore be the 1's complement of x. A 1 would be inserted into the carry terminal in synchronism with the first digit of y and of the 1's complement of x. The sum out of the adder would then be $y-x$. The adder produces a sum within a fraction of a digit time after it receives an input. Thus, there is only a small delay in obtaining the sum of two numbers.

Table I can be consulted to show that the adder of Figure 9 requires 38 crystal diodes, $1^3/_8$ digits of delay, and two active elements. The operation of so many crystal diode circuits in series without amplifiers may be questioned, but it is believed that it will become feasible as diodes and transistors are improved.

Accumulator

A block diagram of an accumulator is shown in Figure 10. The output of the adder is fed back to its input through a w-digit delay line. The output of the delay line is reclocked continuously in the and-circuit in accord with digit pulses from the master clock. The timed pulses then are amplified in a 1-stage amplifier to make up for attenuation in the delay line. Whenever a new accumulation is to be started, an erase signal is fed to the inhibiting circuit. This signal is w-digits long and blocks the output of the delay line from the adder and insures that the new accumulation will start from zero. An accumulator to accumulate sums 48 digits long will require 76 crystals, 48 digits of delay, and eight active elements.

Multiplier

The multiplier that is discussed in the following paragraphs is designed to multiply two positive members, each having W digits, in $W(2W+1)$ digit times. Operating at a megacycle rate, the product of two 48-binary digit numbers would be obtained in 4,656 microseconds. It is believed most efficient to convert all negative members to positive ones before multiplying and to adjust the sign of the product according to the rules of algebraic multiplication. The components to be added to the multiplier to make the sign correction have not been included in this study. The multiplicand x is multiplied by the first (least significant) digit of the multiplier y in the first $(2W+1)$ period and by the W'th digit of y during the W'th $(2W+1)$ period. The W partial products are added together, with each partial product moved over one place with respect to the preceding one as it is accumulated.

The multiplication table in binary arithmetic is very simple. If it is desired to multiply the binary number x by a single binary digit m, x is connected to the input of a switch that is held open if $m=0$, and closed if $m=1$. The output of the switch will be mx. Thus, in multiplying x by the successive digits of the number y, it is only required to pass x through a switch that is opened and closed at the appropriate times according to whether the successive digits of y are zeroes or 1's.

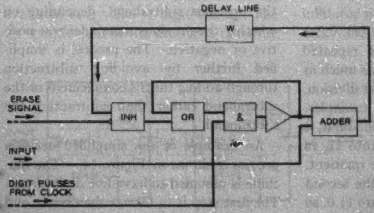


Figure 10 (left). Accumulator

TIME	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
MULTIPLICAND	0	1	1					0	1	1					0	1	1
MULTIPLIER	1	0	1				1	0	1				1	0	1		
START	1							1							1		
SWITCH	←- - -CLOSED- - -→							←- - -OPEN- - -→							←-CLOSED		
ADDEND	0	1	1					0	0	0					0	1	1
AUGEND	0	0	0				0	1	1				0	1	1	0	
SUM	0	1	1				0	1	1	0			0	1	1	1	1

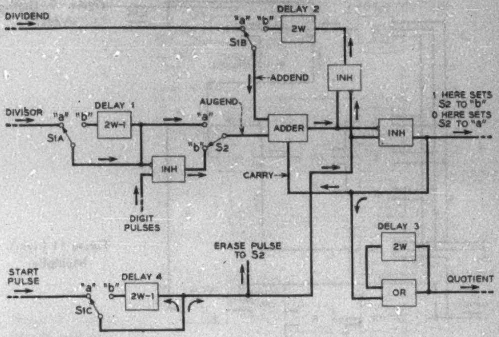
Figure 12 (right). Operation of multiplier

```
DIVIDE 3 (11) BY 8 (1000)

           0.011
     1000 ⌐11.000
           1000
            1000
            1000
```

```
DIVIDE 15 (1111) BY 24 (11000)

            0000.101
     11000 ⌐1111.000
            11000
              11000
              11000
```

Figure 13 (above). Binary division

Figure 15 (right). Divider



The block diagram of a multiplier is shown in Figure 11. The multiplicand x is stored in delay line 1, and the multiplier y is stored in delay line 2. After the first word period, $S1-A$, $S1-B$, and $S1-C$ are thrown to position b and the multiplier, multiplicand, and start pulse recirculate in their delay lines until the multiplication is finished. $DL1$ (delay line 1) storing the multiplicand has one more digit of delay than line $DL2$ in which the multiplier is stored. Therefore, after the first circulation of x through $DL1$, the least significant digit of x leaves $DL1$ just as the second least significant digit of y leaves $DL2$. After the $W-1$'th circulation of x, its least significant digit leaves $DL1$ just as the most significant digit of y leaves $DL2$.

The output of $DL2$ is fed continuously to an and-circuit. Every $2W+1$ digits, an examining pulse is fed to the and-circuit from $DL3$, and if the digit of y coming out of $DL2$ is a 1 at that time, the and-circuit develops a 1 in its output. If, when the examining pulse comes along, the coefficient of y leaving $DL2$ is a zero, the and-circuit will have no output. Because $DL2$ has a delay of $2w$ digits, the and-circuit in effect filters out the successive digits of y, choosing a new digit every $2w+1$ digit times. The examining pulse is also sent to $S2$, where it is the erase signal of the storage cell associated with $S2$. The output of the and-circuit goes to $S2$ as switching instruction and is stored there. Thus, $S2$ stores either a 1 or a zero, according to whether the digit of y examined is a 1 or a zero. The switch is arranged so that when a 1 is stored, $S2$ is closed, and when a zero is stored, it is open. The examining pulse is developed from the start pulse which is received by the multiplier coincident with the first digits of x and y and recirculated in delay line 3.

```
DIVIDE 3 BY 8
WORD LENGTH = 5 PLACES. 5TH PLACE INDICATES SIGN.
3 =                   00011
8 =                   01000
ONES COMPLEMENT =     10111
ADD ONE                   1
                      -----
-8 =                  11000

CHANGE NUMBERS TO 2W-1 OR 9 DIGIT LENGTH.
 3 = 000000011
 8 = 000001000
-8 = 111111000

DIVISION RULE: IF (2W-1)TH (9TH) DIGIT OF REMAINDER
IS ZERO ADD -DIVISOR; ADD +DIVISOR IF DIGIT IS ONE.
                                 0.01100
000001000 ⌐   000000011.00000
ADD -8        111111000
NOTE 9TH DIGIT-->1|11111011
AND ADD           |
+8 WITH SHIFT     |000001000
NOTE 9TH DIGIT--->1|11111110
AND ADD            |
+8 WITH SHIFT      |000001000
NOTE 9TH DIGIT---->0|00000100
AND ADD             |
-8 WITH SHIFT       |111111000
NOTE 9TH DIGIT----->0|00000000
AND ADD              |
-8 WITH SHIFT        |111111000
NOTE 9TH DIGIT------>1|11111000
AND ADD               |
+8 WITH SHIFT         |000001000
NOTE 9TH DIGIT------->1|11111000

THE QUOTIENT IS THE ONES COMPLEMENT OF THE
SUCCESSIVE 9TH DIGITS OF THE REMAINDERS.
```

Figure 14. Simplified division process

A new partial product is obtained at the output of $S2$ every $2w+1$ digit times by the process just described. To accumulate them, an adder is provided which accepts as one input the output of $S2$, and as the other input its own sum output delayed by $2w$ digits. The one digit place difference in the arrival of the two inputs of the adder automatically provides the "shift" feature necessary in adding the partial products. Figure 12 shows a digit by digit account of the operation of a 3-digit multiplier when multiplying 110 by 101.

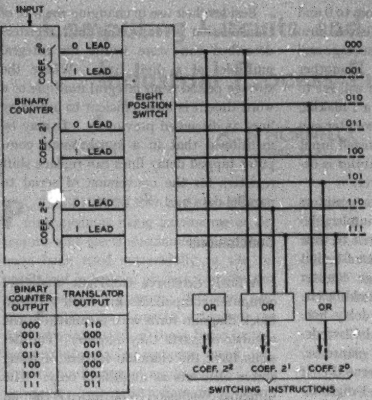
A multiplier based on Figure 12 can be mechanized for 48 digit words with 390 crystals, approximately 390 digits of delay, and 55 active elements.

Divider

The divider operates with repeated subtractions and shift operations much as one would do in pencil and paper division. Figure 13 shows two division examples. In the first one, 3 is divided by 8. First, it is seen that 1,000 will not go into 11, so the first zero is written in the quotient. 100.0 will not go into 11, so the second zero is written. 10.00 will go into 11.0, so a one is written in the quotient and 10.00 is subtracted from 11.00. The remainder is 1.00. The binary point is again shifted in the divisor, giving 1.000 which goes exactly once, and another one is written in the quotient giving an answer of 0.011 or $0 \times 1/2 + 1 \times 1/4 + 1 \times 1/8$ equal to 3/8.

It should be noted that at each step the divisor is either subtracted from the remainder or nothing is subtracted and that a shift to the right is made each time before the divisor is subtracted. In effect, the successive subtrahends are obtained by repeated divisions of the divisor by 2. For machine operation, the awkward part of this process is that it cannot be ascertained whether a subtraction should be made without making it and seeing whether or not the remainder is positive or negative. If the remainder is negative, the subtrahend should be added back. This necessity for sometimes adding back the subtrahend may be avoided. Let x be the subtrahend and R the minuend. Suppose x is subtracted from R and the remainder found to be negative. During the next step the machine should subtract $x/2$ from R, but the machine has lost R by subtracting x from it. R can be regained and $x/2$ subtracted at the same time by adding the factor $x/2$ to the negative remainder $R-x$ which gives the desired result $R-x/2$. Thus, on the successive steps the machine either subtracts half the previous subtrahend or adds half the previous subtrahend, depending on whether the previous remainder was positive or negative. The process is simplified further by avoiding subtraction through adding the 2's complement of the subtrahend rather than subtracting the subtrahend.

An example of the simplified division process is shown in Figure 14. The machine is assumed to have five digit places. The first step is to fill out the divisor and


INPUT
BINARY COUNTER
COEF. 2^0
COEF. 2^1
COEF. 2^2
0 LEAD
1 LEAD
0 LEAD
1 LEAD
0 LEAD
1 LEAD
EIGHT POSITION SWITCH
000
001
010
011
100
101
110
111
OR
OR
OR
COEF. 2^2
COEF. 2^1
COEF. 2^0
SWITCHING INSTRUCTIONS

BINARY COUNTER OUTPUT	TRANSLATOR OUTPUT
000	110
001	000
010	001
011	010
100	000
101	001
111	011

Figure 16 (left). Translator

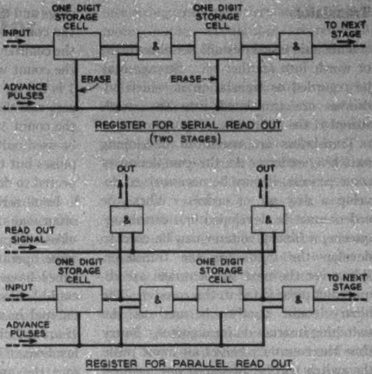

ONE DIGIT STORAGE CELL
ONE DIGIT STORAGE CELL
INPUT
&
&
TO NEXT STAGE
ADVANCE PULSES
ERASE
ERASE
REGISTER FOR SERIAL READ OUT
(TWO STAGES)
OUT
OUT
&
&
READ OUT SIGNAL
ONE DIGIT STORAGE CELL
ONE DIGIT STORAGE CELL
INPUT
&
&
TO NEXT STAGE
ADVANCE PULSES
REGISTER FOR PARALLEL READ OUT

Figure 18 (right). Shift register

dividend to $2w-1$ places, and then to form the complement of the divisor. The divisor complement is first added to the dividend. The $2w-1$'th digit of the sum is noted. In the example, the digit is a 1, indicating a negative result. The machine then writes a zero in the quotient, drops out the $2w-1$'th digit of the previous sum, and adds half the divisor (divisor shifted one place to the right). The $2w-1$'th digit is examined again and is found to be a 1. Another zero is written in the quotient, the $2w-1$'th digit is dropped out, and the divisor is again shifted and added. This time the $2w-1$'th digit is zero, indicating a positive sum. A 1 is written in the quotient, and the complement of the divisor is shifted and added. This process is continued until the 5-digit answer 0.0110 is obtained. Note that the first operation was made with the binary points of the divisor and dividend lined up. In general, it would be necessary to start with the least significant digit of the divisor written under the most significant digit of the dividend. $2w(2w-1)$ digit times are required for division in a general-purpose machine.

The divider block diagram is shown in Figure 15. The divisor is stored in delay line 1 and recirculates there, becoming available every $2w-1$ digit times. The 1's complement of the divisor is taken by the inhibitor circuit fed by $DL1$ (delay line 1). Every $2w-1$ digit times switch $S2$ selects either the 1's complement of the divisor or the divisor itself to go into the adder. The switch selects the 1's complement when it is desired to subtract. However, to obtain subtraction it is necessary to add the 2's complement. This is done by bringing out the carry terminal of the adder, and whenever $S2$ is thrown to select the 1's complement a 1 is sent into the carry terminal and, in effect, makes the augend the 2's complement.

The storage cell of switch $S2$ receives an erase signal every $2w-1$ digit times. At the same time, the switch receives a 1 or a zero from the inhibitor circuit on the adder output. If this inhibitor sends a 1 to the switch, the $2w-1$'th digit of the sum just completed was a zero indicating a positive number and switch $S2$ goes to position b to select the 1's complement of the divisor for the next addition. The output of the inhibitor circuit also is sent to the carry terminal of the adder, so that when $S2$ is at b a 1 will be injected to convert the 1's complement to a 2's complement. Each output of the inhibitor is also a digit of the quotient, for if the $2w-1$'th digit of the sum is a zero, a 1 should be written in the quotient and the inhibitor supplies that 1. It is written in delay line 3, which has a circulation time of $2w$. Since a new quotient digit is received every $2w-1$'th digit time, delay line 3 has as its output the digits of the quotient in reverse order to that in which they are obtained, which puts the least significant digits first as they should be.

The foregoing discussion shows how the augend is obtained for the adder. The addend is obtained from $S1B$. This switch selects the dividend during the first $2w-1$ period, and thereafter selects the previous sum from delay line 2. Note that an inhibitor circuit between the adder and $DL2$ blocks out the $2w-1$'th digit each time to keep the remainder from growing longer than $2w-1$. Since delay line 2 has $2w$ delay and delay line 1 has $2w-1$ delay, the augend arrives one digit early each $2w-1$ period with respect to the addend. This feature supplies the successive divisions by two that are required.

Delay line 4 exists merely to supply the $2w-1$ pulse that is sent to $S2$ and the inhibitor circuit on the adder output. Switch $S1$ is driven to a at the end of the computation period to empty the delay lines of the dividend, divisor, and $2w-1$ pulse preparatory to solving a new problem. After the first $2w-1$ digit times of a new computation, all sections of $S1$ are thrown to b and recirculate their data.

An estimate based on Tables I and II of the parts required by the divider of Figure 16 shows that to handle 48 digit numbers 371 crystal diodes, 390 digits of delay, and 51 active elements would be required. This assumes that the delay lines have their data regenerated every eight digits. If the data are regenerated only every 16 digits, only 36 active elements are required. The divider is seen to be of about the same complexity of the multiplier.

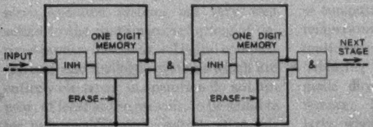

INPUT
INH
ONE DIGIT MEMORY
ERASE
&
INH
ONE DIGIT MEMORY
ERASE
&
NEXT STAGE

Figure 17. Binary counter

Translator

This unit translates one set of numbers or words into another set. Storage may be regarded as translation in which addresses are translated into the words stored at the addresses.

Translators are useful in developing switching orders. At the completion of some process, it may be necessary to develop a new set of orders. When the orders must be developed in a certain sequence, a binary counter can be used to develop the input for the translator. Whenever the next set of orders are required, a pulse is fed to the counter. The binary count number is used as the switching instruction for a switch. Every time the counter receives an input pulse the switch will be advanced one position and will put a pulse on the selected lead. The various output leads of the switch will be connected to a group of or-circuits. One or-circuit will be required for each digit of the translator output. Figure 16 shows a scheme for generating eight sets of 3-digit words. To generate *n* digit words *n* or-circuits would be required. No more than eight leads would be required for any or-circuit. The translator shown develops the output word in parallel. If the or-circuits are fed to taps in a delay line, the output word will be received serially from the output of the delay line.

Binary Counter

Figure 17 shows how each stage of a binary counter can be made from a 1-digit memory cell plus an inhibitor and an and-circuit. In Figure 17, if the count is at 01 and a 1 is sent into the counter, the 1 in the first stage will be admitted to the second stage. The 1 that was in the first stage inhibits the 1 input for the first stage and the stage therefore goes to 0 and the count is 10. If another 1 is sent into the counter, it will enter the first stage and the count will advance to 11. If another 1 is fed in, the first two stages will set to zero and the third will receive a 1 making the count 100. This type of counter can be used only to count synchronized input pulses but that is all that a counter is expected to do in a serial computer.

In a serial machine binary counters often can be replaced by simpler devices. Suppose that it is desired to time some operation and know when 64 digit times have elapsed. A 6-stage counter could be used or a digit pulse could be circulated in a 7- and a 9-digit delay line. If an and-circuit were fed by the two delay lines, 63 digit times would elapse before the and-circuit were operated. If the output were fed through a 1-digit delay line the desired 64 digit delay would be obtained.

Shift Registers

Shift registers are sometimes needed to take data at a low rate, from a magnetic drum, for example, and retransmit it at a rapid rate. A shift register that is loaded serially may be read out in parallel on *W* leads or in series on one lead. Each stage of such registers can be built around a 1-digit memory cell. In Figure 18 are illustrated two stages of each kind of register. The numbers in each cell are erased and advanced to the next cell whenever an advance pulse is received.

Advance pulses will be received every time a new digit is to be loaded. When a number is in the register, the advance pulses can be stopped and the number can be stored until it is required. In the case of the parallel-read-out registers, extra and-circuits are used at each stage. These are fed by the read-out signal.

Besides their use in changing the rate of transmission of data, the shift registers described can store words for integral multiples of a digit time. Where the storage period is an integral multiple of a word time it is more efficient to use delay lines as discussed previously. It may be mentioned that in a synchronous computer tapped delay lines can replace shift registers for the conversion of serial to parallel data and vice versa.

Conclusions

A fairly extensive catalogue of digital computer components has been given in block diagram form with estimates of the number of parts they require. The designs have the common feature of using active elements as amplifiers only. The principal conclusion to be drawn is that an all-semiconductor computer can be built with diodes and transistors. This paper has been written as a contribution towards the development of such a computer.

References

1. HIGH SPEED COMPUTING DEVICES (book), **Engineering Research Associates.** McGraw-Hill Book Company, New York, N. Y., 1950.

2. SEAC, THE NATIONAL BUREAU OF STANDARDS EASTERN AUTOMATIC COMPUTER. *Technical News Bulletin*, National Bureau of Standards (Washington, D. C.), volume 34, number 9, September 1950, pages 1–8.

3. THE TRANSISTOR AS A DIGITAL COMPUTER COMPONENT, **J. H. Felker.** *AIEE Special Publication S-44:* REVIEW OF ELECTRONIC DIGITAL COMPUTERS, February 1952, pages 105–09.

4. RECTIFIER NETWORKS FOR MULTIPOSITION SWITCHING, **D. R. Brown, H. Rochester.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 37, number 2, February 1949, pages 139–47.

5. DIODE COINCIDENCE AND MIXING CIRCUITS IN DIGITAL COMPUTATION, **Tung Chang Chen.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 38, number 5, May 1950, pages 511–14.

No Discussion

Computing Circuits and Devices for Industrial Process Functions

A. J. HORNFECK
MEMBER AIEE

POTENTIOMETER and bridge-balance systems using galvanometer detectors and electromechanical amplifiers have been used successfully in the recording of industrial process variables for many years. Electronic detection and amplification has greatly extended the field of application of these basic instruments by increasing response speed, sensitivity, and range of application.[1,2] The versatility and standardization of design made possible by the electronic servo amplifier has permitted the utilization of the accurate null-type metering system to the continuous measurement and control of practically all process variables which can be converted to electrical functions. The electronically operated null-balance recorder is admirably suited to perform as an analogue calculator, and circuits and devices are available which permit the computation of derived process variables including sums, differences, products, ratios, powers, roots, and trigonometric functions. The electric computing systems offer numerous advantages over equivalent mechanical methods. These include:

1. Transmitters (also called transducers or pickups) whose outputs are to be included in the calculation can be located remotely from one another in their most favorable location relative to process. In this way the electric system offers the additional advantage of short-range telemetering for centralized process control.
2. Complete electronically operated computing systems are built up using the basic standard components of the simple measuring systems. Mechanical computers generally require special design of linkages, cams, and gears.
3. Functions can be computed accurately over wide ranges and with relatively straightforward calibration procedure. Mechanical computers involving complex linkages and cams are generally limited and require special calibrations for each application.

Analogue-type computers using the servo operated systems are very well suited to the continuous recording of derived process functions. Digital computers and also analogue systems which are entirely electronic in operation do not appear, at least at the present, to be suitable for industrial process control. For this service the requirements are sturdiness of design, the necessity for remote indication and recording of results, long life, reliability, and ease of service. The fundamental mathematical accuracy of the more complex computing machines usually is not required since the accuracy almost always is limited by the primary measuring element or the transducer which converts the physical quantity to an electrical function.

Some of the applications of servo actuated computers of the automatic null-balance type are:

1. Summation of fluid flows in chemical and other type industrial processes.
2. Computation of ratios of related factors as, for example, fuel-air ratio in combustion control and control of the ratio of chemicals to obtain a desired solution in a process.
3. Calculation of heat released or absorbed in heat exchangers and refrigeration systems.
4. Calculation of over-all plant efficiencies as, for example,, that of a steam-electric generating station.
5. Accurate compensation of measurements as, for example, automatic correction of gas flow meters for wide range variations in the density of the gas.
6. Automatic and continuous solution of theoretical and empirical equations relating a number of process variables such as pressure, temperature, pH, conductivity, and flow to a final product output.

The Elementary System

Figure 1 is a diagram of the automatic null-balance system of measurement and shows in block form the basic components of the system. The wide application of the null-balance system to the measurement and control of process variables is illustrated. In some applications it is possible to perform simple computations such as addition, subtraction, and division directly in the basic measuring circuit. For example, temperature difference can be computed by simply incorporating two temperature detectors in a bridge or potentiometer circuit in such a way that the balancing position of the follow-up is made directly proportional to the difference. In other applications, particularly where the energy output of the measuring element is very small or where more complex calculations are required, retransmission from a primary instrument is employed.

If the pickup is a primary measuring device such as a thermocouple, resistance thermometer, or photocell, the physical variable is converted directly into an electrical output such as current, voltage, or resistance. The miscellaneous or "quality" variables such as pH, vacuum, conductivity, and gas analysis are best measured by electric or electronic primary elements. Consequently, the output of these elements can sometimes be used directly in computers without the intermediate mechanical link. On the other hand, quantities such as pressure, fluid flow, and liquid level are converted most frequently into mechanical factors such as force or motion by means of a mechanical metering element or link which may be a float, diaphragm, bellows, liquid column, or bourdon tube. This metering element in turn actuates a position transmitter or pickup device whose electrical output is proportional to the physical variable. For many straightforward algebraic functions this output can be combined with the output of similar elements to perform the desired computation. However, in some applications the motion of the metering element bears a nonlinear relation to the process variable, as for example: fluid flow is proportional to the square root of differential pressure drop across an orifice or flow nozzle. If such outputs are to be combined for totalizing and similar functions, the nonlinear quantity must be converted to an electrical factor proportional to the desired quantity. In these applications the function is first extracted in a primary null balance type of receiver and retransmitted as an electrical quantity proportional to the process variable.

The principles discussed are illustrated by a British-thermal-unit calculating system shown diagrammatically in Figure 2. It is desired to compute the rate at which heat is absorbed by the fluid flowing through the heat exchanger. Two temperature-sensing elements T_1 and T_2 are placed in the inlet and outlet side of the heat exchanger and connected into an automatically balanced differential bridge or potentiometer circuit so that the mo-

---

Paper 52-191, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted March 3, 1952; made available for printing May 7, 1952.

A. J. HORNFECK is with the Bailey Meter Company, Cleveland, Ohio.

tion of the motor driven follow-up is proportional to temperature difference. The balancing motor also actuates a retransmitter which may be a resistance slide-wire. Flow is measured by means of a mechanical-type element responsive to the differential pressure drop across an orifice or other type of primary element in the pipe. This pickup device converts differential pressure into a telemetered electric variable and the motion or position of the follow-up in the flow receiver is responsive to this differential. Since the flow is proportional to the square root of the differential, the measured function is converted to flow in the receiver by means of a properly shaped cam operating a retransmitter whose output is made proportional to the flow. The output of the flow retransmitter is then combined with the output of the temperature-difference receiver in a multiplying circuit in which the receiver moves proportional to the product of the retransmitted variables to produce an indication or recording of the rate of British-thermal-unit absorption. A more detailed description of a typical computing system is presented in a later section of this paper.

Pickups and Their Circuits

It is not the purpose of this paper to describe in detail the wide variety of primary metering elements, transmitting, and pickup devices as well as the basic measuring circuits of the null type frequently used in potentiometer and bridge-balance instruments. These are either well known or are adequately described elsewhere in the literature. Primary measuring elements may be grouped into two general categories; namely, the self-generating type in which a small amount of energy from the process is converted into an output voltage, and the transducer type in which the measured variable influences either in direct proportion or by some known relation the electrical characteristics of the pickup device. Examples of the first class are thermocouples, photocells of the barrier layer type, and pH electrodes. Examples of the transducer type are resistance temperature detectors, resistance wire strain gages, electrolytic conductivity cells, and heated filaments for gas analysis by thermal conductivity.

Of more specific interest in the computing systems are the position pickup devices which convert position or motion into an electrical output. The following types of position transmitters and function generators are used in the analogue computing systems.

1. The adjustable core transformer, sometimes called the differential transformer or linear induction potentiometer.
2. The resistance slide-wire used either as a variable resistance or as a rheostat potentiometer.
3. Induction potentiometers for converting angle of rotation to a proportional output voltage.
4. Sine-cosine electrical resolvers for converting angle of rotation to electric voltages proportional to sine and cosine functions.

The first two types of simple position pickups are used most extensively in the type of process function computers covered by this paper and will be described in detail. Both of these devices may be used either as transmitters directly coupled to mechanically operated metering devices or as retransmitters for generating voltages proportional to a function of the original variable. They are also used as servo-motor operated follow-ups in intermediate and final recording receivers.

Adjustable Core Transformer

Figure 3 illustrates the principles and application of the adjustable core transformer which is used extensively in the measurement of factors which are convertible to position or motion. The magnetic core of the transmitter is moved by the mechanical metering element which in some cases actually may be a complete meter which indicates or records the variable and in others only a basic link between the primary detecting element and the transmitting unit. Figures 3(A) and 3(B) show the structure and principle of operation. The magnetic flux linkages between the primary winding and each of the two secondaries is determined by the magnetic core position. Consequently the voltages induced in the secondary windings vary linearly and differentially in opposite directions in accordance with the following equations.

$$E_1 = E_0 + kx$$
$$E_2 = E_0 - kx \qquad (1)$$

where

E_0 = voltage when the core is in the exact center position
x = per-unit motion of core = actual motion divided by total motion
k = total change in voltage per winding at the extreme of core position

In the ratio-balance circuit of Figure 3(C) the secondaries are connected so that their voltages aid. At balance the

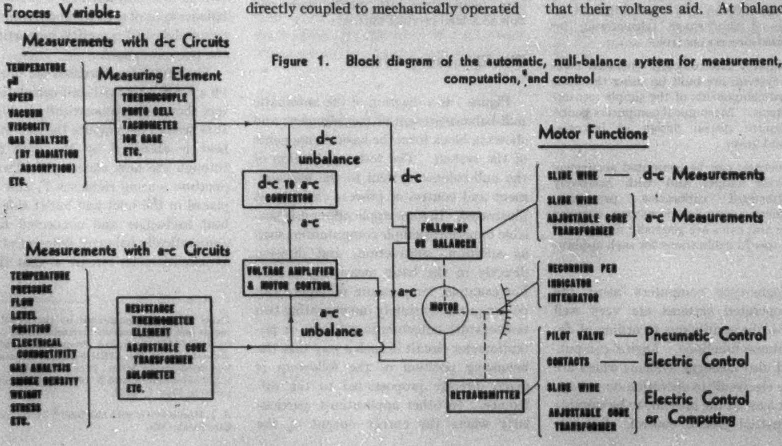
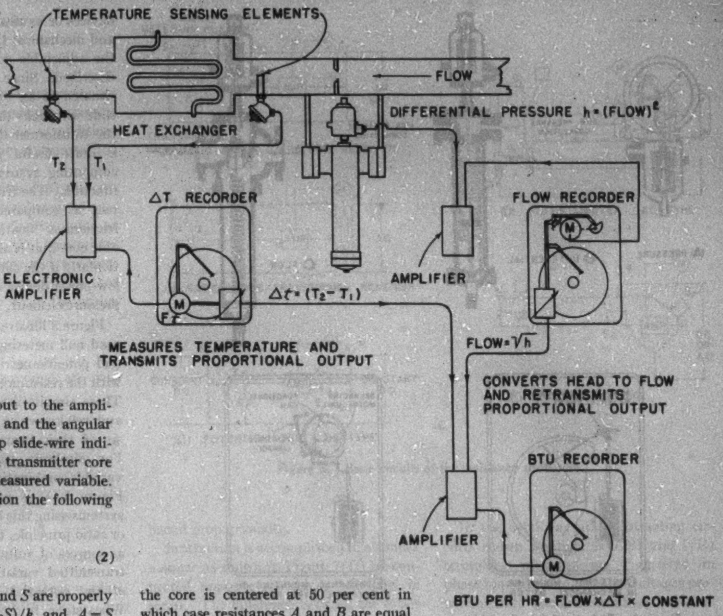

Process Variables
Measurements with d-c Circuits
TEMPERATURE
pH
SPEED
VACUUM
VISCOSITY
GAS ANALYSIS (BY RADIATION ABSORPTION)
ETC.
Measuring Element
THERMOCOUPLE
PHOTO CELL
TACHOMETER
ION GAGE
ETC.
d-c unbalance
d-c TO a-c CONVERTOR
a-c
d-c
Measurements with a-c Circuits
TEMPERATURE
PRESSURE
FLOW
LEVEL
POSITION
ELECTRICAL CONDUCTIVITY
GAS ANALYSIS
SMOKE DENSITY
WEIGHT
STRESS
ETC.
RESISTANCE THERMOMETER ELEMENT
ADJUSTABLE CORE TRANSFORMER
BOLOMETER
ETC.
VOLTAGE AMPLIFIER & MOTOR CONTROL
a-c unbalance
a-c
FOLLOW-UP OR BALANCER
MOTOR
RETRANSMITTER
Motor Functions
SLIDE WIRE — d-c Measurements
SLIDE WIRE
ADJUSTABLE CORE TRANSFORMER
a-c Measurements
RECORDING PEN
INDICATOR
INTEGRATOR
PILOT VALVE — Pneumatic Control
CONTACTS — Electric Control
SLIDE WIRE
ADJUSTABLE CORE TRANSFORMER
Electric Control or Computing

Figure 1. Block diagram of the automatic, null-balance system for measurement, computation, and control

Figure 2. Diagram of a typical process computing system—British-thermal-unit calculator


TEMPERATURE SENSING ELEMENTS
FLOW
HEAT EXCHANGER
DIFFERENTIAL PRESSURE h = (FLOW)²
T₂
T₁
ΔT RECORDER
FLOW RECORDER
M
ELECTRONIC AMPLIFIER
AMPLIFIER
M
F
Δt = (T₂ − T₁)
MEASURES TEMPERATURE AND TRANSMITS PROPORTIONAL OUTPUT
FLOW = √h
CONVERTS HEAD TO FLOW AND RETRANSMITS PROPORTIONAL OUTPUT
BTU RECORDER
AMPLIFIER
M
BTU PER HR = FLOW × ΔT × CONSTANT

fundamental voltage input to the amplifier is 0 or a minimum, and the angular position of the follow-up slide-wire indicates the position of the transmitter core and consequently the measured variable. For this balance condition the following equation applies.

$$\frac{E_0+kx}{E_0-kx}=\frac{B+MS}{A+(1-M)S} \qquad (2)$$

If the values of A, B and S are properly chosen so that $B=(E_0S)/k$ and $A=S(E_0/k-1)$, the slide-wire rotation M will equal the core position x. In this analysis it is assumed that the magnetic core is in the center of the coil when the transmitted variable is zero. In many applications the core is centered at 50 per cent in which case resistances A and B are equal or $A=B=S[E_0/(k-1/2)]$.

In the potentiometric circuit shown in Figure 3(D) the secondary windings of the transmitter are connected in phase opposition so that the total voltage is a proportional function of the core position. The identical transformer unit in the receiver is positioned by the motor and the system is balanced when the transmitter and the receiver core positions are ex-

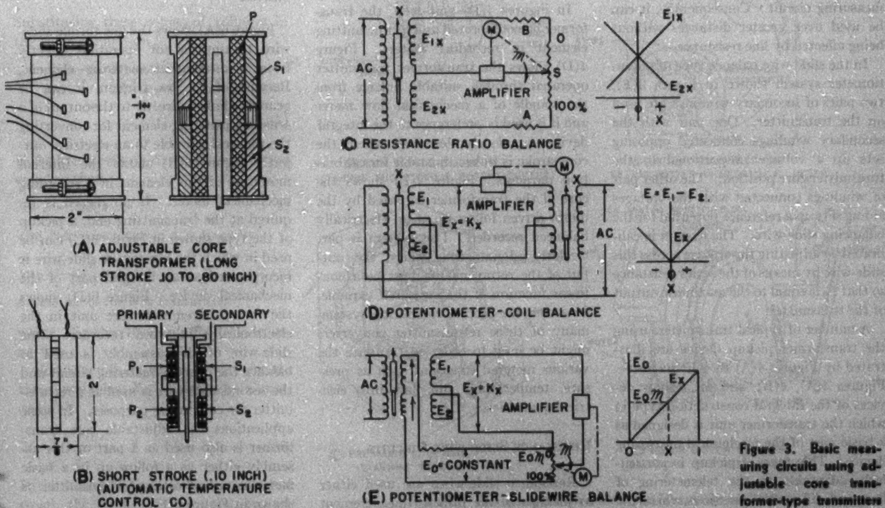

P
S₁
S₂
3 1/2"
2"
(A) ADJUSTABLE CORE TRANSFORMER (LONG STROKE .10 TO .80 INCH)
PRIMARY
SECONDARY
P₁
S₁
P₂
S₂
2"
7/8"
(B) SHORT STROKE (.10 INCH) (AUTOMATIC TEMPERATURE CONTROL CO)
X
AC
E₁x
M
B
O
S
AMPLIFIER
E₂x
100%
A
E₁x
E₂x
− 0 +
X
(C) RESISTANCE RATIO BALANCE
M
X
X
E₁
AMPLIFIER
Eₓ = Kₓ
E₂
AC
E = E₁ − E₂
Eₓ
− 0 +
X
(D) POTENTIOMETER-COIL BALANCE
AC
E₁
Eₓ = Kₓ
E₂
AMPLIFIER
E₀ = CONSTANT
E₀M
M
100%
M
E₀
Eₓ
E₀M
0
X
1
(E) POTENTIOMETER-SLIDEWIRE BALANCE

Figure 3. Basic measuring circuits using adjustable core transformer-type transmitters

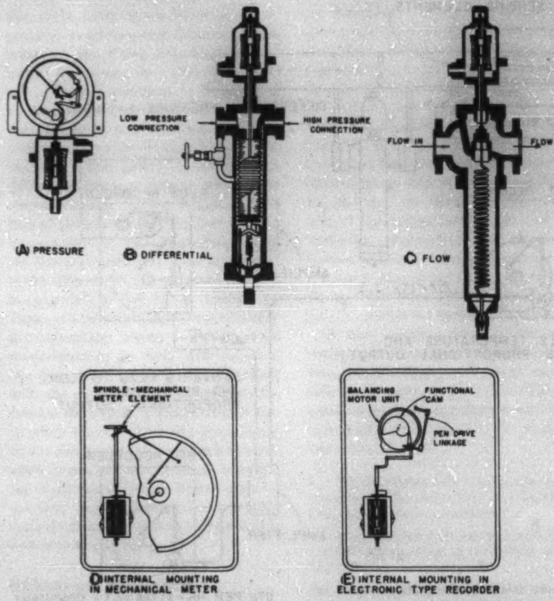


Figure 4. Pickups of the adjustable core transformer type for specific applications

actly the same. This system has several advantages over the one shown in Figure 3(C), the most important being the fact that no current flows in the leads of the measuring circuit. Consequently, it can be used over greater distances without being affected by line resistance.

In the slide-wire balance type of potentiometer system shown in Figure 3(E), two pairs of secondary windings are used on the transmitter. One pair with the secondary windings connected opposing sets up a voltage proportional to the transmitter core position. The other pair of windings connected with the voltages aiding sets up a reference potential for the balancing slide-wire. The system is calibrated by adjusting the voltage across this slide-wire by means of the series resistance so that E_0 is equal to the maximum output of the transmitter.

A number of typical transmitters using the transformer pickup device are illustrated by Figures 4(A) to 4(E) inclusive. Figures 4(A), 4(B), and 4(C) show devices of the integral construction type in which the transformer unit is designed as a basic part of the mechanical measuring device. This type of pickup is particularly adaptable to the telemetering of flow and differential pressures where the moving element of the basic meter device including the transformer core can be enclosed entirely in a high-pressure chamber.

In Figures 4(D) and 4(E), the transformer is incorporated as a retransmitting element in recording meters. Figure 4(D) shows the transformer transmitter operated through suitable linkage from the spindle of a mechanical flow meter and it is used in preference to the integral devices previously described where the core stroke is to be adjustable for calibration purposes. Figure 4(E) shows the core of the transformer actuated by the motor-driven follow-up of an electrically balanced recorder. This system is particularly advantageous where the output of the retransmitter must be a nonlinear function of the measured variable. In a complex process computing system, many of these retransmitter converters might be used to order to translate the various metered functions such as pressure, temperature, and flow into electrical quantities.

Resistance Slide-wire Circuits

Resistance slide-wires are used either as pickups in the primary measurement devices or as retransmitters in electronic and mechanical type meters, similarly to the adjustable core transformer already described. Since the transmitter may be situated in an unattended location, the slide-wire lacks the inherent sturdiness of the transformer device. Nevertheless, it is quite useful particularly where the computing system requires a resistance function. The friction of the sliding contact as compared with the essentially frictionless "floating action" of the movable core unit is also a disadvantage, particularly if the slide-wire is actuated by a low-power device such as a Bourdon tube pressure element.

Figure 5 illustrates the most commonly used null metering circuits of the bridge and potentiometric type which are used with the resistance slide-wire transmitter. These circuits are basically very simple and well known so that no detailed analysis of their characteristics is required. For algebraic calculations of process variable combinations the circuit shown in Figure 5(D) is particularly useful. In systems using this basic a-c potentiometer or ratio principle, the slide-wires function as sources of voltage proportional to the transmitted variables. The advantages of the potential balance circuits over Wheatstone bridge circuits perhaps are fairly obvious, but the most important ones are:

1. Elimination of contact resistance errors.
2. Relative ease of calibration and adjustment to take care of wide ranges in magnitudes.

Figure 6(A) shows the low-torque slide-wire transmitter for operation directly from a mechanical metering element. Because of its low friction, it can be geared or linked directly to the output of a sensitive metering element for converting the physical variable to an electrical output. Figure 6(B) shows the internal mounting of this element in a recording mechanical meter. If no record is required at the transmitting end, a pickup of the type shown in Figure 6(C) can be used in which the low-torque slide-wire is incorporated as an integral part of the mechanical device. Figure 6(D) shows the motor-driven slide-wire unit in the electronically balanced recorder. One slide-wire of this assembly is used to balance the primary metering circuit and the second slide-wire is used as a retransmitter for computing purposes. In some applications the adjustable core transformer is also used as a part of this assembly either as a follow-up in a basic metering circuit or as a retransmitter as shown in Figure 5(E).

Basic Computing Circuits

TRANSFORMER TYPE

Figure 7 shows the elementary circuits using the transformer transmitter for the simple arithmetic functions of addition, subtraction, and division of transmitted variables.

In the circuit of Figure 7(A) the per-unit rotation of the motor-driven balancing slide-wire is equal to the total or average motion of all the transmitter cores. The lower or zero end windings all are connected in series aiding so that the total voltage of one leg of the bridge is given by

$$E_2 = E_{x2} + E_{y2} + E_{z2} + \ldots E_{n2} \tag{3}$$

Similarly, all the top windings are connected in series aiding so that

$$E_1 = E_{x1} + E_{y1} + E_{z1} + \ldots E_{n1} \tag{4}$$

at balance

$$\frac{E_1}{E_2} = \frac{E_{x1} + \ldots E_{n1}}{E_{x2} + \ldots E_{n2}} = \frac{B+MS}{A+S-MS} \tag{5}$$

where

E_{n1} = voltage output of the top winding of the *N*th transmitter

E_{n2} = voltage output of the bottom winding of the *N*th transmitter

If the cores are centered at 0.5 of maximum stroke

$$\begin{aligned} E_{x1} &= [E_0 - (k/2)] + kx \\ E_{n1} &= [E_0 - (k/2)] + kn \\ E_{x2} &= [E_0 + (k/2)] - kx \\ E_{n2} &= [E_0 + (k/2)] - kn \end{aligned} \tag{6}$$

Substituting these values of voltages in the balance equation 5 yields

$$E_1 = \frac{N[E_0 - (k/2)] + k(x + \ldots n)}{N[E_0 + (k/2)] - k(x + \ldots n)} = \frac{B+MS}{A+S-MS} \tag{7}$$

If the fixed resistors *A* and *B* are adjusted equal so that

$$A = B = S[E_0/k - (^1/_2)] \tag{8}$$

The slide-wire rotation *m* is given by

$$M = \frac{x + \ldots n}{N} \tag{9}$$

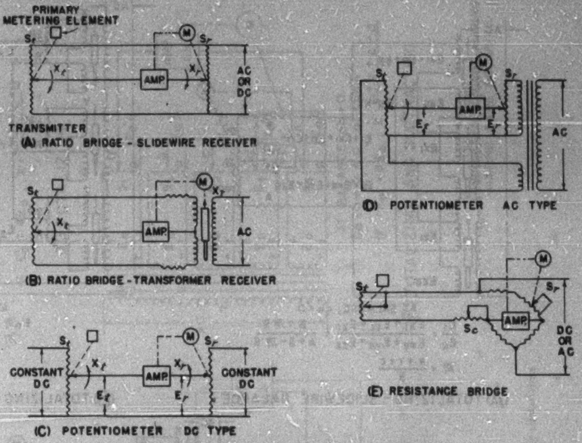
n = motion of the *N*th transmitter

where N = the total number of transmitters.

For example, in Figure 7(A) there are three variables *X*, *Y*, and *Z*. Consequently

$$M = \frac{X+Y+Z}{3} \tag{10}$$

If the ranges of the transmitters are not equal, the strokes of the cores are adjusted proportionally.


PRIMARY METERING ELEMENT
S_t
S_r
M
X_t
X_r
AMP
AC OR DC
TRANSMITTER
AC
CONSTANT DC
E_t
E_r
S_c
DC OR AC

(A) RATIO BRIDGE - SLIDEWIRE RECEIVER

(B) RATIO BRIDGE - TRANSFORMER RECEIVER

(C) POTENTIOMETER DC TYPE

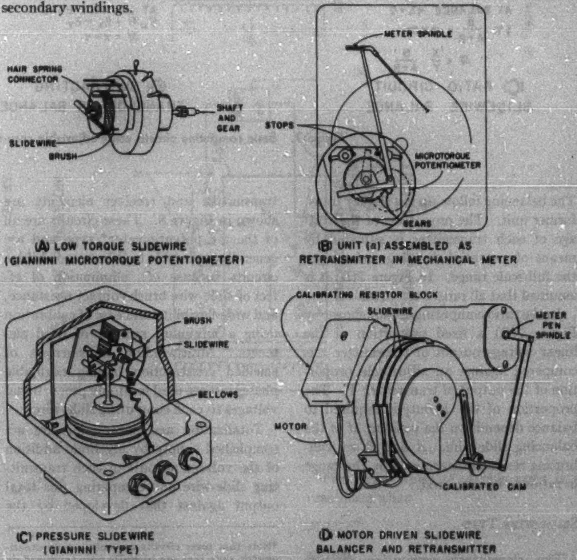
(D) POTENTIOMETER AC TYPE

(E) RESISTANCE BRIDGE

Figure 5. Basic circuits of the resistance slide-wire

Subtraction is accomplished in a similar manner as shown in Figure 7(D) by connected opposite secondary windings in series. The difference between the totals of two groups of variables can be accomplished by properly interconnecting the secondary windings.

In the totalizing and subtracting circuits shown by Figures 7(B) and 7(D) secondary windings are connected in phase opposition to produce voltages proportional to each transmitted variable.


HAIR SPRING CONNECTOR
SHAFT AND GEAR
SLIDEWIRE
BRUSH
METER SPINDLE
STOPS
MICROTORQUE POTENTIOMETER
GEARS
BRUSH
SLIDEWIRE
BELLOWS
CALIBRATING RESISTOR BLOCK
SLIDEWIRE
METER PEN SPINDLE
MOTOR
CALIBRATED CAM

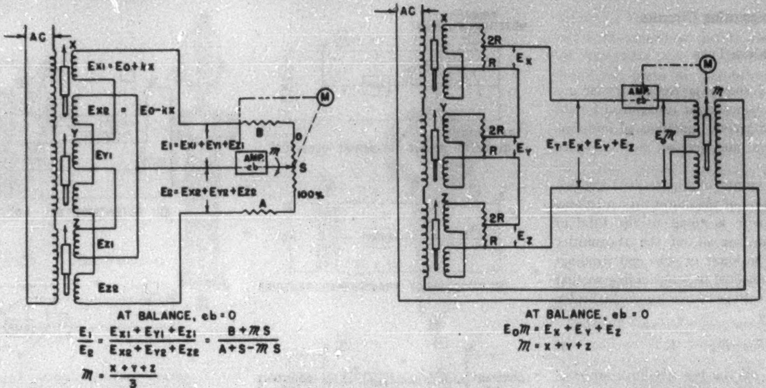
(A) LOW TORQUE SLIDEWIRE (GIANINNI MICROTORQUE POTENTIOMETER)

(B) UNIT (a) ASSEMBLED AS RETRANSMITTER IN MECHANICAL METER

(C) PRESSURE SLIDEWIRE (GIANINNI TYPE)

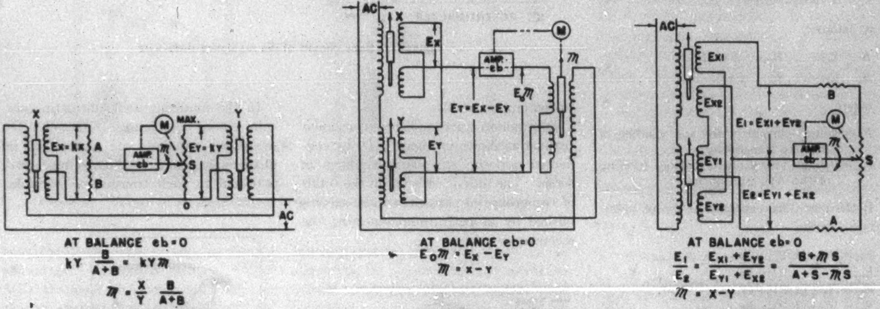
(D) MOTOR DRIVEN SLIDEWIRE BALANCER AND RETRANSMITTER

Figure 6. Pickups of the slide-wire type for specific applications


AC
X
Y
Z
Exı = Eo + kx
Ex2 = Eo − kx
Eyı
Ezı
Ez2
Eı = Exı + Eyı + Ezı
E2 = Ex2 + Ey2 + Ez2
AMP. eb
M
B
A
O
S
100%
AT BALANCE, eb = 0
E1/E2 = (Ex1 + Ey1 + Ez1)/(Ex2 + Ey2 + Ez2) = (B + M S)/(A + S − M S)
M = (x + y + z)/3
2R
R
Ex
Ey
Ez
ET = Ex + Ey + Ez
EoM
AT BALANCE, eb = 0
EoM = Ex + Ey + Ez
M = x + y + z

(A) TOTALIZING – SLIDEWIRE BALANCE

(B) TOTALIZING – TRANSFORMER BALANCE


X
Y
Ex − kx
Ey − ky
A
B
MAX.
M
AMP. eb
S
O
AC
AT BALANCE eb = 0
kY B/(A+B) = kYM
M = X/Y · B/(A+B)
Ex
Ey
ET = Ex − Ey
EoM
AT BALANCE eb = 0
EoM = Ex − Ey
M = x − y
Exı
Ex2
Eyı
Ey2
Eı = Exı + Ey2
E2 = Eyı + Ex2
AT BALANCE eb = 0
E1/E2 = (Ex1 + Ey2)/(Ey1 + Ex2) = (B + M S)/(A + S − M S)
M = X − Y

(C) RATIO CIRCUIT SLIDEWIRE BALANCE

(D) SUBTRACTING TRANSFORMER BALANCE

(E) SUBTRACTING SLIDEWIRE BALANCE

Figure 7. Basic computing circuits using adjustable core transformers

The balancing follow-up is a similar transformer unit. The proportion of the voltage of each transmitter is adjusted by means of potential dividing resistors to the full-scale range. In Figure 7(B) it is assumed that all ranges are equal.

In the ratio computing circuit shown by Figure 7(C) a fixed proportion of the linear voltage output of transmitter x is compared against an adjustable proportion of the output of transmitter Y. The proportion of the Y output required to balance depends on the position M of the balancing slide-wire. A and B are calibrating resistors determined by the range in ratios to be measured.

Slide-wire Type

The elementary computing circuits most frequently used with slide-wire transmitter and receiver elements are shown in Figure 8. These circuits are all of the a-c potentiometer* type and are generally preferable to Wheatstone bridge circuits because of: elimination of effect of slide-wire brush contact resistance, and wide flexibility and ease of calibration using a minimum number of fixed elements. Multiwinding transformers of shielded construction having negligible phase shift are used to supply proportional voltages to each computing slide-wire.

Totalization and subtraction are accomplished simply by algebraic addition of the voltage outputs of each transmitting slide-wire and comparing this total output against that developed by the balancing slide-wire in the receiver. Multiplication and division of transmitted functions are accomplished as shown in Figures 8(A) and 8(B) by connecting the computing slide-wires in cascade. To avoid the error produced by the loading effect of the second slide-wire, it is necessary that its resistance be high relative to the first. The deviation from linearity of the voltage output of the second slide is determined by the following analysis.

The voltage across the second slide-wire is given by

$$e = \frac{E_0 A M}{A + M - M^2} \tag{11}$$

E_0 = supply voltage
M = per unit rotation of first slide-wire
$A = S_2/S_1$ where S_2 and S_1 are the resistances of the first and second slide-wire respectively.

---

*Note that these circuits are only pseudopotentiometric since they are basically ratio types and do not measure a-c voltages except in terms of a percentage or portion of the supply voltage.

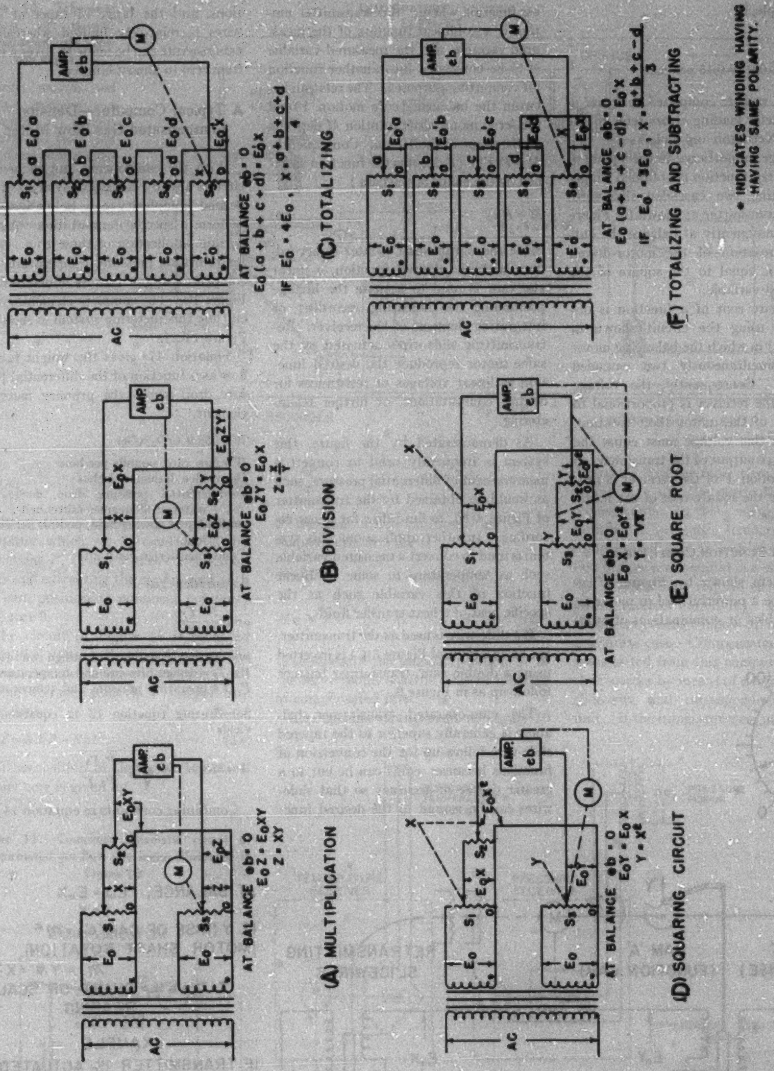

AC
E₀
S₁
S₂
AMP eb
M
(A) MULTIPLICATION
AT BALANCE eb = 0
E₀Z = E₀XY
Z = XY
(B) DIVISION
AT BALANCE eb = 0
E₀ZY = E₀X
Z = X/Y
(C) TOTALIZING
AT BALANCE eb = 0
E₀(a + b + c + d) = E₀'X
IF E₀' = 4E₀ , X = (a+b+c+d)/4
(D) SQUARING CIRCUIT
AT BALANCE eb = 0
E₀Y = E₀X
Y = X²
(E) SQUARE ROOT
AT BALANCE eb = 0
E₀X = E₀Y²
Y = √X
(F) TOTALIZING AND SUBTRACTING
AT BALANCE eb = 0
E₀(a + b + c − d) = E₀'X
IF E₀' = 3E₀ , X = (a+b+c−d)/3
* INDICATES WINDING HAVING HAVING SAME POLARITY.

Figure 8. Basic computing circuits using slide-wires.

The theoretical linear voltage across the second slide-wire assuming its resistance to be infinitely large is

$$e_l = E_0 M \tag{12}$$

The deviation from linearity is given by

$$\Delta e = \frac{e_l - e}{E_0} \tag{13}$$

or

$$\Delta e = \frac{M^2(1-M)}{A+M(1-M)} \tag{14}$$

For large values of A this error is a maximum when

$M \cong 0.67$

so that

$$\Delta e \text{ max} \cong 0.15/A \tag{15}$$

For example, if

$A = 100$

$\Delta e \max = 0.0015 = 0.15$ per cent

For a rather complete analysis of potentiometer loading errors refer to the paper by Nettleton and Dole.[3]

A voltage proportional to the square of the measured function is obtained simply by actuating two cascaded slide-wires from the transmitter as shown by Figure 8(D). Consequently at balance the output and rotation of the motor-driven slide-wire is equal to the square of the transmitted variable.

The square root of a function is obtained by using the circuit shown in Figure 8(E) in which the balancing motor operates simultaneously two cascaded slide-wires. Consequently, the voltage output of the receiver is proportional to the square of the motor shaft rotation. At balance this voltage must equal the linear voltage output of the transmitter so that the motion Y of the receiver is proportional to the square root of the measured variable.

Nonlinear Function Computer

The system shown by Figure 9 has proved to be a powerful tool in metering single variables or combinations of process function where: the transmitter motion is a nonlinear function of the measured variable, and the measured variable is to be converted into another function for computing purposes. The relation between the balancing core motion Y and the servo motor shaft rotation M depends on the shape of the Y cam. Consequently $M = F(Y)$. (F denotes the function specified by the shape of the cam.)

$$M = F(x) \tag{16}$$

Since the motor shaft rotates in accordance with the desired function, a linear rise cam is used to actuate the linakge positioning the indicating, recording, or integrating elements of the receiver. Retransmitting slide-wires actuated by the same motor reproduce the desired function as linear voltages or resistances for control, computations, or further telemetering.

As demonstrated by the figure, this system is frequently used to convert a measurement of differential pressure, such as would be obtained by the transmitter of Figure 4(B), to fluid flow for linear recording. In other applications this system is used to convert a measured variable such as temperature to some nonlinear function of this variable such as the specific heat of a heat transfer fluid.

If a slide-wire is used as the transmitter, the basic circuit of Figure 3(C) is inverted using a double cam, transformer balance follow-up as in Figure 9.

The cam-operated transformer balancer is generally superior to the tapered slide-wire follow-up for the conversion of functions because: cams can be cut to a greater degree of accuracy so that slide-wires can be wound to the desired functions, and the range of taper of slide-wires is relatively limited whereas the cam rise rate can be made to cover a range from zero to almost infinity.

A Typical Computer—Density Compensated Gas Flow Meter

With the components and elementary circuits already described, systems of any desired complexity can be designed to perform a specific computation. An important application of these to a system for recording pressure, temperature, and compensated gas flow is illustrated by Figure 10. The complete elementary circuit for this metering system is shown in Figure 11.

Equation 17[4] gives the weight rate of flow as a function of the differential pressure drop across the primary metering element.

$$W = 358.9\, CfD_2^2\sqrt{h\rho} \tag{17}$$

W = flow rate, pounds per hour
D_2 = orifice diameter, inches
h = differential pressure drop, inches of water at 68 degrees fahrenheit
ρ = density of flowing fluid, pounds per cubic foot
C and f are design constants

For an ideal gas

$$\rho = \rho s\frac{P}{Ps}\frac{Ts}{T} \tag{18}$$

ρs = density at standard or design conditions
Ps, Ts = design pressure and temperature
P, T = operating pressure and temperature

Substituting equation 13 in equation 12 yields

$$W = 358.9CfD_2^2\sqrt{hPs\frac{P\,Ts}{Ps\,T}} \tag{19}$$

Combining constants in equation 14

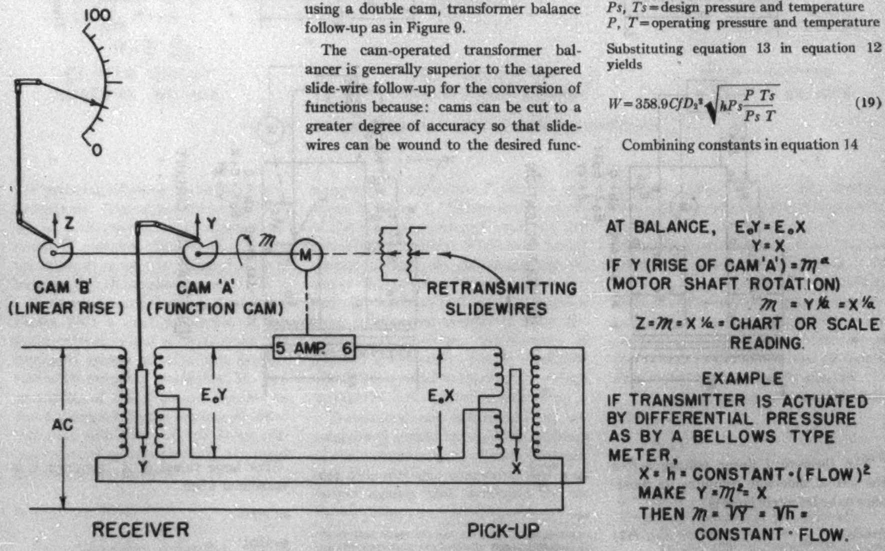
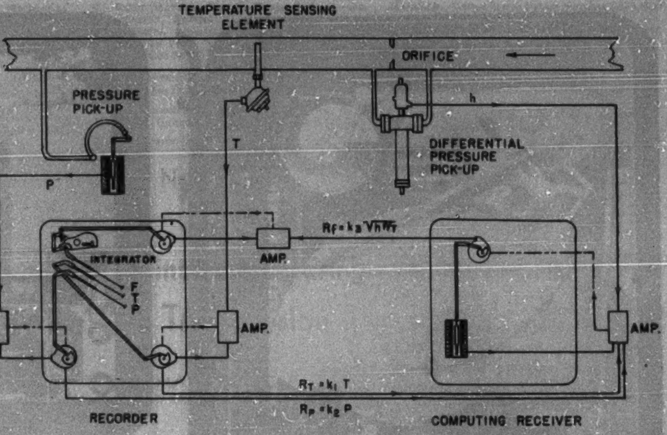

100
0
Z
Y
m
M
CAM 'B' (LINEAR RISE)
CAM 'A' (FUNCTION CAM)
RETRANSMITTING SLIDEWIRES
5 AMP. 6
AC
E_0Y
E_0X
Y
X
RECEIVER
PICK-UP
AT BALANCE, $E_0Y = E_0X$
$Y = X$
IF Y (RISE OF CAM 'A') $= m^a$
(MOTOR SHAFT ROTATION)
$m = Y^{1/a} = X^{1/a}$
$Z = m = X^{1/a}$ = CHART OR SCALE READING.
EXAMPLE
IF TRANSMITTER IS ACTUATED BY DIFFERENTIAL PRESSURE AS BY A BELLOWS TYPE METER,
$X = h$ = CONSTANT · (FLOW)^2
MAKE $Y = m^2 = X$
THEN $m = \sqrt{Y} = \sqrt{h}$ = CONSTANT · FLOW.

Figure 9. Nonlinear function computer

Figure 10. Schematic diagram of density compensated gas flow rate meter

Recorder records and transmits pressure and temperature, records compensated flow, and integrates compensated flow

Computing receiver computes and transmits compensated flow


TEMPERATURE SENSING ELEMENT
ORIFICE
PRESSURE PICK-UP
T
h
DIFFERENTIAL PRESSURE PICK-UP
P
$R_F = k_3\sqrt{hW_T}$
INTEGRATOR
AMP.
F
T
P
AMP.
AMP.
AMP.
$R_T = k_1 T$
$R_P = k_2 P$
RECORDER
COMPUTING RECEIVER

$$W = K_3\frac{\sqrt{hP}}{T} \qquad (20)$$

The system is essentially an analogue computer which solves equation 20 by converting P, T, and h to electrical quantities and extracting the square root with the cam positioning nonlinear computer of Figure 9.

The circuit of Figure 11 is balanced when the voltage developed across the pressure compensating slide-wire is equal to the voltage across the temperature slide-wire.

$$KYT = KXP = KhP \qquad (21)$$

and the position of the balancing transformer core is given by

$$Y = \frac{hP}{T}$$

If the cam which positions the core is shaped so that its rise is given by

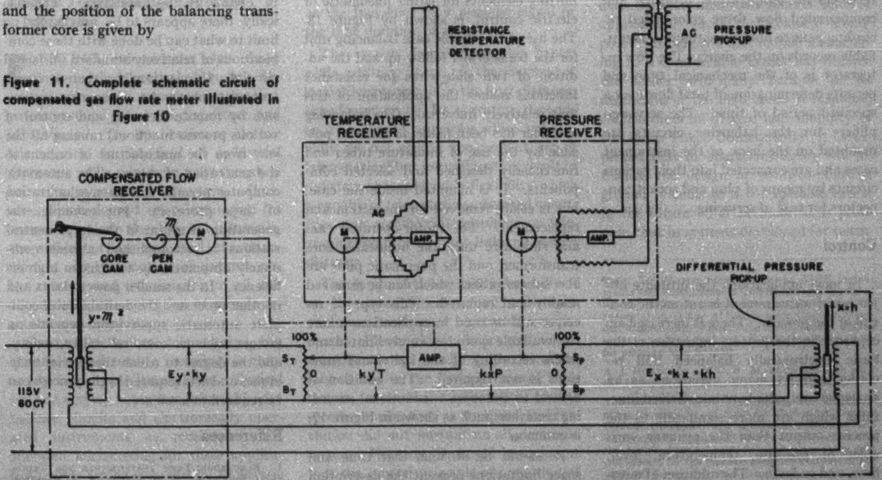
$$Y = M^2$$

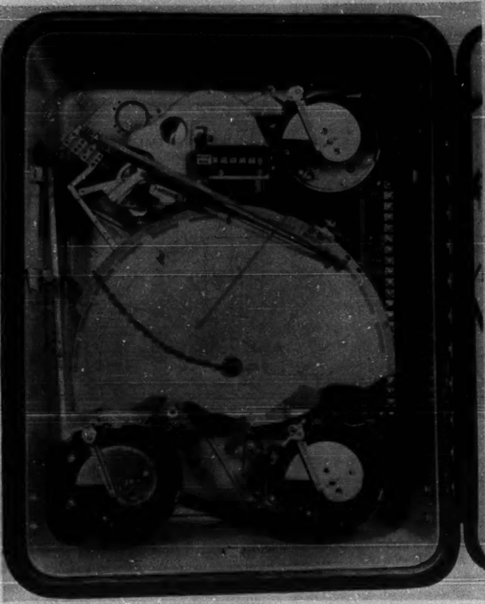
$$M = \sqrt{Y} = \frac{\sqrt{hP}}{T} = W \qquad (22)$$

Consequently the angular position M of the servo motor shaft is a linear function of compensated flow

The specific arrangement of Figure 10 illustrates how the basic elements can be arranged to suit the particular metering requirements. In this system correlating records of pressure, temperature, and compensated flow as well as an integration of flow are required in a single meter. Space limitations require that the computing receiver employing the cam positioned transformer follow-up be placed in a separate case. Compensated flow is retransmitted from this computing unit to the recorder by means of a retransmitting slide-wire and corresponding receiving unit. If the integrator were not required,

Figure 11. Complete schematic circuit of compensated gas flow rate meter illustrated in Figure 10


RESISTANCE TEMPERATURE DETECTOR
AC
PRESSURE PICK-UP
TEMPERATURE RECEIVER
PRESSURE RECEIVER
COMPENSATED FLOW RECEIVER
AC
M
AMP.
M
AMP.
$\mathcal{M}$
M
CORE CAM
PEN CAM
DIFFERENTIAL PRESSURE PICK-UP
$y = \mathcal{M}^2$
$x = h$
100%
100%
S_T
O
B_T
AMP.
S_P
O
B_P
115V 60CY
$E_y = ky$
kyT
kxP
$E_x = kx = kh$



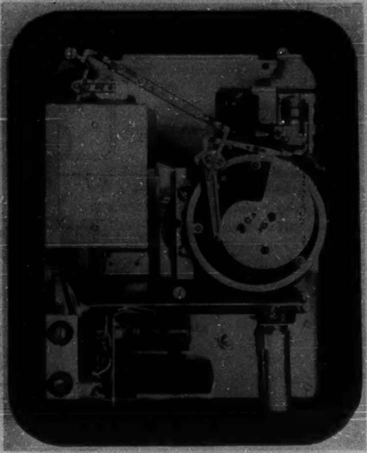


Figure 12 (left). A 3-element receiver for recording pressure, temperature, and compensated flow rate with integration of flow

Figure 13 (above). Nonrecording receiver for computing, controlling, and retransmitting

the entire metering could be accomplished by a single, multifunction type of recording receiver.

Figure 12 shows the 3-pen recording receiver presented diagrammatically in Figure 11. The motor-driven slide-wire receivers for temperature, pressure, and compensated flow drive color-coded recording pens to obtain continuous, identifiable records on the chart. The flow integrator is of the mechanical type and permits determination of total flow over a specified period of time. The servoamplifiers for the balancing circuits are mounted on the back of the instrument cases and are connected into their various circuits by means of plug and socket connectors for ease of servicing.

Control

In most applications the ultimate objective of measurement is automatic control of the process. There is increased acceptance in the process industries of the basic electronically balanced null recorder and its many variations as an analogue computer to obtain derived functions which are more significant to the process output than the primary variables of pressure, temperature, level, flow, and so forth. The addition of automatic control to these telemeter computers is a further step in the complete mechanization of the process plant.

A newly developed nonrecording type of computing receiver, incorporating in a relatively small instrument case retransmitting elements for either pneumatic or electric control, is shown by Figure 13. The use of the double cam balancing unit for the transformer follow-up and the addition of two slide-wires for resistance functions makes the application of this unit relatively universal. The balancing amplifier has been made as small as possible by the use of miniature tubes and functionally designed and selected components. It is mounted inside the case, but is easily removed for inspection and replacement. The motor-operated cam and slide-wire unit, the adjustable core transformer, and the pneumatic pilot are also subassemblies which can be removed readily and replaced. This type of receiver will be used in applications where the available space is severely limited and where recording of the individual functions is not required. The addition of control to the more conventional recording receivers, such as shown in Figure 12, is common.

It should be obvious that time and space limitations prevent the description of the more elaborate and complete computing and control systems which would be used to meter and control a complex process. The number of combinations of metering and control which can be incorporated even in one case, such as described in Figure 12, is quite large. Theoretically, there appears to be practically no limit to what can be done with these combinations of relatively standard units and circuits. Unquestionably, there is an increased tendency for automatic supervision by remote metering and control of various process functions, ranging all the way from the manufacture of cement to the generation of steam. The automatic computer permits further mechanization of these processes. For example, the generation of steam in the large central stations is accomplished almost completely automatically to achieve high efficiency. In the smaller power plants and in other processes, the desirability of complete automatic supervision depends on factors such as cost, reliability, safety, and the degree to which the special techniques and arts required in the process can be reduced to a science.

References

1. ELECTRONIC-TYPE INSTRUMENTS FOR INDUSTRIAL PROCESSES, P. S. Dickey, A. J. Hornfeck,

Transactions, American Society of Mechanical Engineers (New York, N. Y.), volume 67, 1945, pages 393–98.

2. ELECTRONIC RECORDING INSTRUMENTS. *Electrical Engineering*, volume 66, number 1, January 1947, pages 36–44.

3. REDUCING POTENTIOMETER LOADING ERROR, L. A. Nettleton, F. E. Dole. *Review of Scientific Instruments*, American Institute of Physics (New York, N. Y.), volume 18, number 5, May 1947, pages 332–41.

4. FLUID METERS—Part I—THEIR THEORY AND APPLICATION, Special Research Committee on Fluid Meters. American Society of Mechanical Engineers (New York, N. Y.), fourth edition, August 1937, equation 99, page 49.

5. TELEMETRIC CONTROL SYSTEM, A. J. Hornfeck. United States Patent Number 2,330,427.

6. MEASURING AND CONTROLLING SYSTEM, A. J. Hornfeck. United States Patent Number 2,420,539.

7. MEASURING AND CONTROLLING SYSTEM, A. J. Hornfeck. United States Patent Number 2,439,891.

8. MEASURING SYSTEM, A. J. Hornfeck. United States Patent Number 2,406,221.

9. ELECTROMAGNETIC MOTION RESPONSIVE DEVICE, W. D. MacGeorge. United States Patent Number 2,427,866.

10. SOME ASPECTS OF ELECTRICAL COMPUTING, J. Bell. *Electronic Engineering* (London, England), July 1951, pages 254–69.

No Discussion

A Long-Distance Multipoint Telemetering System Using Teletype Transmission

A. J. HORNFECK
MEMBER AIEE

G. R. MARKOW
ASSOCIATE MEMBER AIEE

Synopsis: This paper describes a point-to-point telemetering system using pulse width modulation and sequential scanning for transmission of several measured variables over a single circuit by teletype. The system, which is essentially electromechanical rather than electronic, operates on the basic principle whereby the magnitude of a measured variable is converted into a pulse of proportional time width. The system was developed specifically for remote monitoring of pumping stations on a pipeline to permit automatic operation from a central dispatch office. In the design of the equipment emphasis was placed on requirements of reliability, sturdiness, simplicity, and ease of servicing rather than on factors such as high speed and excessive sensitivity.

Electric Telemetering in Process Industries

ELECTRIC telemetering has been used in the electrical industry for many years as a means of operating power systems, generating stations, and substations.[1] The application of electric telemetering to the field of industrial processes, while relatively more recent, is becoming increasingly important. Processes which depend on telemetering for operation include the chemical industries, the making of iron and steel, the transmission and distribution of natural gas, the refining and transmission of oil and oil products, the operation of atomic reactors, and many others.

Short-range or local telemetering, frequently called remote metering, is commonly used inside large plants. Typical systems employ potentiometric and bridge balance circuits and electronically operated instruments as receivers. Air-operated telemetering for these applications offers strong competition to electric systems up to distances of several thousand feet, particularly for the measurement and control of the nonelectrical variables such as pressure, temperature, level, and flow. The electric systems have the advantages of greater range and almost instantaneous response, but since the accuracy and resolution is affected by the attenuation of the transmitted signal, the longest distances normally used are approximately 5 miles for a-c systems and up to 50 miles for d-c potentiometric systems. In applications where the system involves distances of the order of 100 or even 1,000 miles as, for example, an oil pipeline system, measured quantities are converted into some factor of time such as the frequency of a periodic signal or the time duration of a pulse.

The system described in this paper, while utilizing some of the devices of local telemetering, is of the long-distance, time-factor type. Its operation depends on a number of principles which individually are not new in the present art of telemetering[2] but in combination result in a system of considerable flexibility which is capable of telemetering measured quantities over commercial telephone or telegraph circuits or by radio link any desired distance.

The design of this system is basically of an electromechanical nature utilizing a minimum number of electronic components. Electronic amplifiers and relays actually are used only in the local receivers and in the "sweep balance" detectors, since the requirements of the specific application as to speed and resolution did not warrant an all-electronic system. However, it should be apparent that the general principle of operation of the system is not restricted to the specific design which was developed for the pipeline application.

This long-distance telemetering system was developed as part of a co-operative project to provide completely automatic operation for pumping stations on a products pipeline of the Shell Oil Company running from Wood River, Ill., to Toledo and Columbus, Ohio. These pumping stations are controlled from the central dispatching office located in New York City.

The terminology employed in this paper is taken both from the point-to-point and the mobile or radio telemetering fields and the choice of a term used in describing a principal function is selected on the basis of its accuracy and conciseness. It is important to note that basic principles of operation are common to both groups of telemetering and the differences in equipment exist principally because of the special techniques that are utilized in the two general systems. These differences in techniques depend primarily on the differences in requirements which exist between these two general fields of application. From the technical standpoint it is believed that mobile telemetering to a considerable extent has become the leader principally because of its utilization of the high speed of response, compactness of design, and mobility of the radio link. Consequently, it is expected that mobile telemetering will influence the design of point-to-point or ground systems to a considerable extent in the future. Furthermore, it is probable that in systems developed for telem-

---

Paper 52-197, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted January 11, 1952; made available for printing May 9, 1952.

A. J. HORNFECK and G. R. MARKOW are both with the Bailey Meter Company, Cleveland, Ohio.

The authors wish to acknowledge the assistance given by engineers of Shell Oil Company, General Electric Company, and American Telephone and Telegraph Company, all of whom collaborated in the design of the telemetering and supervisory control system for the pipeline pumping stations of the Shell Oil Company's East Line.

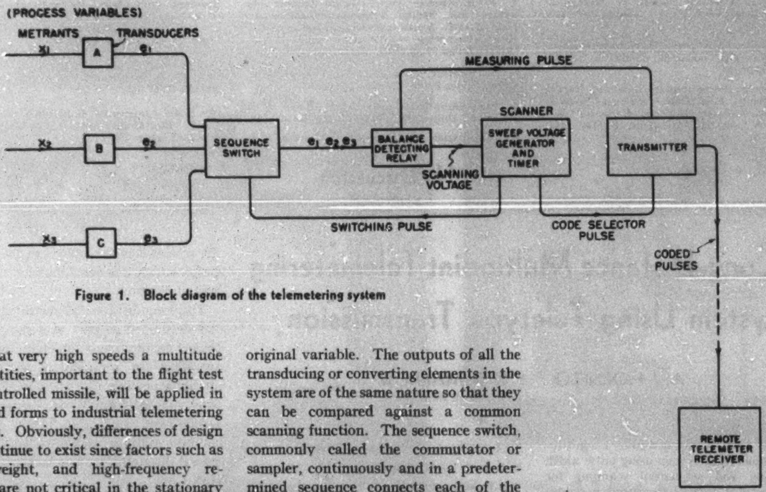


Figure 1. Block diagram of the telemetering system

etering at very high speeds a multitude of quantities, important to the flight test of a controlled missile, will be applied in modified forms to industrial telemetering systems. Obviously, differences of design will continue to exist since factors such as size, weight, and high-frequency response are not critical in the stationary systems, whereas long life, reliability, ease of service by nonelectronic personnel, and other factors are important in the industrial field.

The Basic System

The telemetering system described in this paper is essentially a combination of local telemetering units of the servo-actuated position type[3] coupled to a long-distance system which operates on a general principle of what the mobile telemetering engineers call a "time division, pulse coded multiplexing system." Actually, it is also what the point-to-point telemetering engineers designate as an impulse duration multiplex or multipoint system. In this scheme each measurand (measured quantity) is sampled in a sequence by means of a scanner and then transmitted to the final receiving station or stations as a series of coded characters. Each channel of information is identified by the code selected and each measurement of the various quantities is sampled in sequence and transmitted as a complete series of coded characters by converting the width of the pulse as determined by the scanner into a proportional number of coded characters. Figure 1 is the basic diagram of the system. The measurands may be electrical but in process industries are more frequently nonelectrical quantities such as pressure, flow, level, and so forth. Each of the quantities is converted either directly or through several stages of local telemetering into an electrical quantity proportional to the original variable. The outputs of all the transducing or converting elements in the system are of the same nature so that they can be compared against a common scanning function. The sequence switch, commonly called the commutator or sampler, continuously and in a predetermined sequence connects each of the transducers into the scanning circuit where a comparison is made between their outputs and the time varying output voltage of the scanner. If the output of the pickup is a voltage directly proportional to the measurand, the output of the scanner during the rise period is made a linear function of time. As will be explained in detail later, the function of the scanner is to convert the magnitude of the output of each transducer into a proportional time pulse.

The proportional time interval is represented by the time of dwell of an electric contact in this system but also may be represented by the period of conduction of an electron tube. This pulse may be transmitted without further modification over a wire link. If a radio link is used, an amplitude- or frequency-modulated carrier is required.

The transmitted quantity may be identified at the receiver by its chronological position in a train of impulses such as might be indicated on a cathode-ray oscillograph. In the specific system, developed for the pipeline telemetering, identification is accomplished by converting each pulse into a group of coded teletype characters. The coding is accomplished by synchronizing the teletype equipment with the sequence switch.

Description of the Function of the Elements in the System

Pickups

There are many primary detectors or pickups in use today whose purpose is to convert the measurand to some electrical quantity which can be measured conveniently. Figure 2 shows a number of the transducers, which convert the process variables such as pressure and flow to proportional alternating voltages by means of differential transformers, whose output voltages are proportional to their core position.[3]

The pickups shown by this figure are employed where the basic measuring elements are mechanical devices such as a bellows, Bourdon tubes, floats, and so forth. This type of mechanical link produces a position or motion proportional to the measurand which then is converted into an electrical variable by means of the adjustable core transformer coupled to the mechanical device. It will be noted that these systems represent two and sometimes more stages for the conversion of the measurand into an electrical variable. On the other hand pickups such as thermocouples and resistance thermometers are primary detectors which perform the conversion in a single stage.

Figures 2(A), (B), and (C) show pickups whereby pressure, differential pressure, and flow respectively may be measured and converted to proportional voltages. For variables such as pressure and level the arrangement shown in Figure 2(D) is suitable if a record is desired at the transmitter. In the case of nonlinear variables a secondary pickup is employed as indicated in Figure 2(E). Here the primary detector is balanced by a motor-

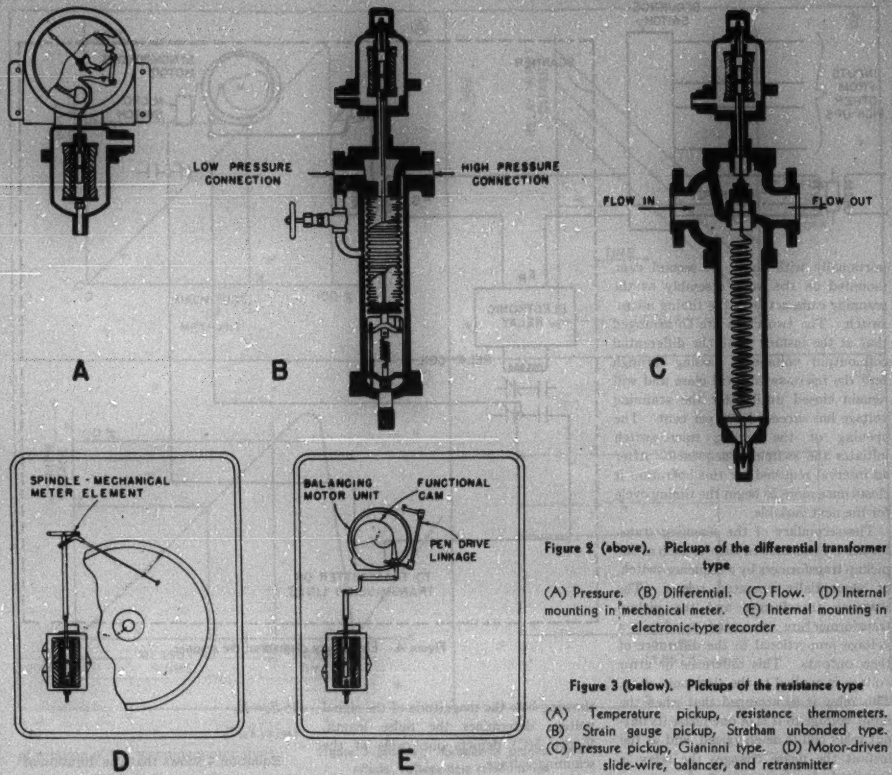

LOW PRESSURE CONNECTION
HIGH PRESSURE CONNECTION
FLOW IN
FLOW OUT
A
B
C
SPINDLE - MECHANICAL METER ELEMENT
BALANCING MOTOR UNIT
FUNCTIONAL CAM
PEN DRIVE LINKAGE
D
E

Figure 2 (above). Pickups of the differential transformer type

(A) Pressure. (B) Differential. (C) Flow. (D) Internal mounting in mechanical meter. (E) Internal mounting in electronic-type recorder

driven slide-wire unit and retransmitted to the differential transformer by means of the correct nonlinear function cam to give an output voltage which varies linearly with the measurand.

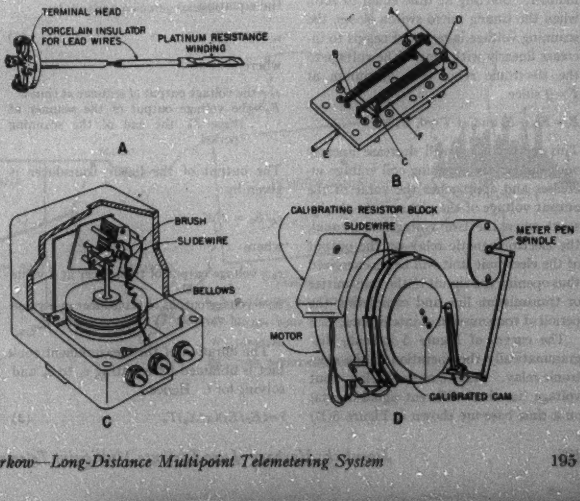
Figure 3 shows several pickups of the resistance type. The resistance thermometer and the strain gauge are primary detectors in which the resistance varies directly with the measurand. The pressure pickup shown in Figure 3(C) functions by positioning the slider of a low-torque potentiometer. Figure 3(D) shows the manner in which a retransmitting slide-wire is actuated by the motor of a self-balancing recorder.

The Scanner

The scanner shown in Figure 4, consists of a synchronous motor driving a linear rise cam which actuates the core of the differential transformer in such a way that its output voltage varies pro-

Figure 3 (below). Pickups of the resistance type

(A) Temperature pickup, resistance thermometers. (B) Strain gauge pickup, Stratham unbonded type. (C) Pressure pickup, Gianinni type. (D) Motor-driven slide-wire, balancer, and retransmitter


TERMINAL HEAD
PORCELAIN INSULATOR FOR LEAD WIRES
PLATINUM RESISTANCE WINDING
A
A
D
B
E
C
F
B
BRUSH
SLIDEWIRE
BELLOWS
C
CALIBRATING RESISTOR BLOCK
SLIDEWIRE
METER PEN SPINDLE
MOTOR
CALIBRATED CAM
D

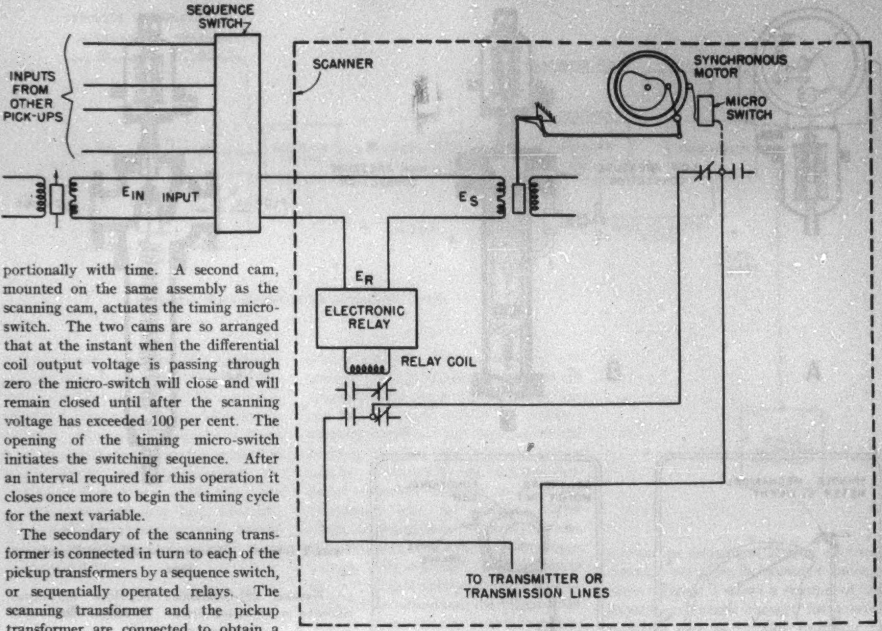


Figure 4. Elementary diagram of the scanner

portionally with time. A second cam, mounted on the same assembly as the scanning cam, actuates the timing micro-switch. The two cams are so arranged that at the instant when the differential coil output voltage is passing through zero the micro-switch will close and will remain closed until after the scanning voltage has exceeded 100 per cent. The opening of the timing micro-switch initiates the switching sequence. After an interval required for this operation it closes once more to begin the timing cycle for the next variable.

The secondary of the scanning transformer is connected in turn to each of the pickup transformers by a sequence switch, or sequentially operated relays. The scanning transformer and the pickup transformer are connected to obtain a voltage proportional to the difference of their outputs. This difference or error voltage is applied to the electronic relay. This relay is so arranged that when the input voltage to it is of one phase its output relay coil is de-energized, and at the instant the phase changes the relay is energized, closing its contacts.

The scanner operates in the following manner. Starting at time equal to zero, when the timing micro-switch closes, the scanning voltage is zero and begins to increase linearly with time. The voltage to the electronic relay is a maximum at $T=0$ since

$E_r=E_{in}-E_s$ and at $T=0$, $E_s=0$

This relay voltage will decrease linearly with time as the scanning coil voltage increases and approaches the value of the output voltage of the pickup. At the instant when the two voltages are equal, the electromagnetic relay on the output of the electronic unit will be de-energized thus opening the circuit to the transmitter or transmission line and completing the period of transmission of intelligence.

The curves of Figure 5 indicate diagrammatically the operation of the electronic relay. Three curves of relay input voltage (for three different signal inputs) on a time base are shown in Figure 5(B) showing how the magnitude of the signal voltage determines the pulse length. Figure 5(C) depicts one cycle of the scanning voltage.

Over the measurement interval the reference voltage of the scanner is given by the equation

$$e_s/E_s=t/T_0 \qquad (1)$$

where

e_s = the voltage output of scanner at time t
E_s = the voltage output of the scanner at time T_0, the end of the scanning period

The output of the linear transducer is given by

$$e_x/E_x = x/X_0 \qquad (2)$$

where

e_x = voltage output of transducer at a value of variable x
E_x = voltage output of transducer at a value of variable X_0 (100 per cent)

The duration of the measurement contact is obtained by equating e_s to e_x and solving for t. Hence

$$t=(E_x/E_s)(x/X_0)T_0 \qquad (3)$$

If $E_s=E_x$

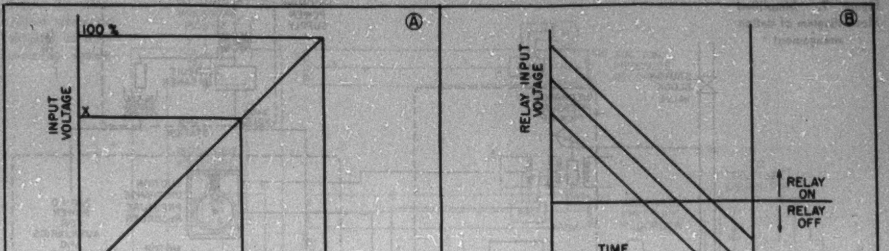
$$t=(x/X_0)T_0 \qquad (4)$$

Equation 4 shows that the duration of the measurement contact is directly proportional to the measurand x.

If the output of the transducer is a nonlinear function of the measurand, this function can be extracted by using a servo-actuated receiver to operate through a cam, the core of a retransmitting transformer. For example, the voltage output of the differential transmitter shown by Figure 2(B) is proportional to the square of flow. In this case the retransmitter cam has a square root rise characteristic so that the output of the retransmitter transformer will be proportional to flow. If, however, all of the quantities to be scanned have the same nonlinear characteristic, the cam of the scanning unit may be shaped so that its rise is a similar nonlinear function of motor shaft rotation. For example, in the case of the differential-type flow meter

$$e_x=E_x(x/X_0)^2 \qquad (5)$$

If the rise of the scanning cam is made proportional to the square of the angle



of rotation of the scanning motor shaft

$$e_s = E_s(t/T_0)^2 \qquad (6)$$

Consequently, if $E_s = E_x$, the electronic relay will be energized for the period t obtained by equating 5 and 6.

$$t = (x/X_0)T_0 \qquad (7)$$

as in the linear system.

Figure 5 (above). Graphs of process magnitude, voltage-time relationships

Figure 6 (below). Map indicating locations of pipeline pumping stations

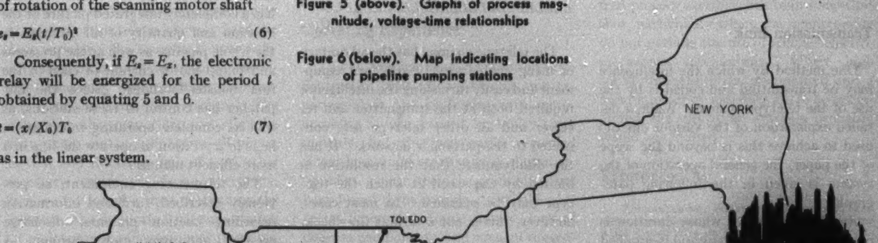
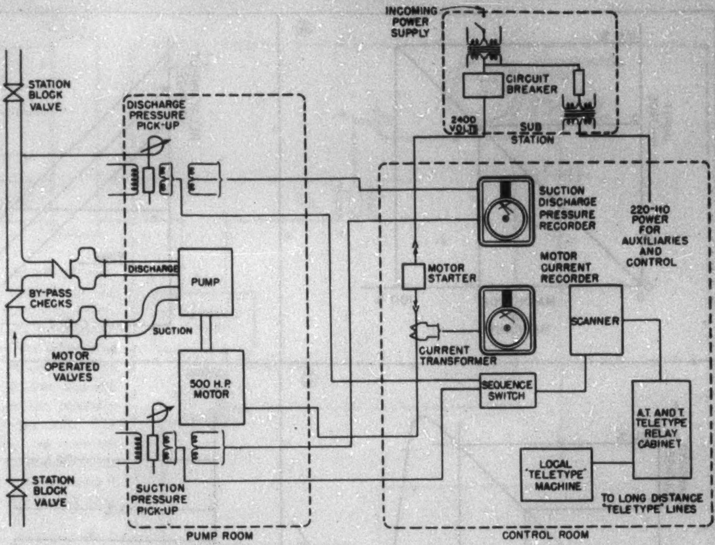

NEW YORK
PENNSYLVANIA
NEW YORK CITY
ILLINOIS
INDIANA
OHIO
TOLEDO
LIMA
COLUMBUS
NORTH SALEM
EFFINGHAM
DENNISON
NEW DOUGLAS
WOOD RIVER REFINERY
LEGEND
AUTOMATIC PUMP STATION
MAJOR PUMP STATION
TERMINAL
CHIEF DISPATCHER'S OFFICE

Figure 7. Simplified block diagram of station arrangement


INCOMING POWER SUPPLY
CIRCUIT BREAKER
2400 VOLTS
SUB STATION
STATION BLOCK VALVE
DISCHARGE PRESSURE PICK-UP
SUCTION DISCHARGE PRESSURE RECORDER
220-110 POWER FOR AUXILIARIES AND CONTROL
DISCHARGE
PUMP
MOTOR STARTER
MOTOR CURRENT RECORDER
BY-PASS CHECKS
SUCTION
SCANNER
MOTOR OPERATED VALVES
CURRENT TRANSFORMER
500 H.P. MOTOR
SEQUENCE SWITCH
A.T. AND T. TELETYPE RELAY CABINET
LOCAL "TELETYPE" MACHINE
STATION BLOCK VALVE
SUCTION PRESSURE PICK-UP
TO LONG DISTANCE "TELETYPE" LINES
PUMP ROOM
CONTROL ROOM

It is entirely feasible to use a rotating slide-wire as the scanner device, and for high-speed telemetering a calibrated electronic generator could be used. For the requirements of the pipeline telemetering system the cam-operated scanner has a number of basic advantages such as elimination of sliding contacts, ruggedness, and ease of calibration.

Transmission Link

One method by which the intelligence may be transmitted and coded is by the use of the teletype system. While a detailed explanation of the various circuits used to achieve this is beyond the scope of the paper, the general operation of the system is given in the following paragraphs.

An electric contact whose duration is proportional to the measurand is supplied by the electronic balance detecting relay of the scanner to the teletype transmitting equipment. The teletype apparatus sets up coding letters for each variable, such as is shown next, using the letter S for suction pressure, T for temperature, F for flow, L for level, and so forth. The equipment starts printing at a constant rate the appropriate letter at the beginning of the scanner timing cycle and stops when the scanner contacts open. Thus, the number of letters in any one sequence is proportional to the variable. At the end of the transmission of intelligence, a code letter such as PQ is sent to indicate the location of the transmitter in cases where there may be similar intelligence channels from other remote stations.

SSSSSSSSSSSSSSSSSSSS PQ
TTTTTTTTTTTTTTTTTTTTTTTTT
PQ
FFFFFFFFFFFFFF PQ
LLLLLLLLLLLLLLLLLLLLLLLLLLLLL-
LLLLLLLLLLLLLL PQ

The teletype channel has the advantage of using comparatively standard equipment and easily furnishing the intelligence required both at the transmitter and receiver and all other teletype sets connected to the particular network. It has the disadvantage that the resolution is limited by the speed at which the teletype may be operated. In most cases, however, this is not a serious drawback.

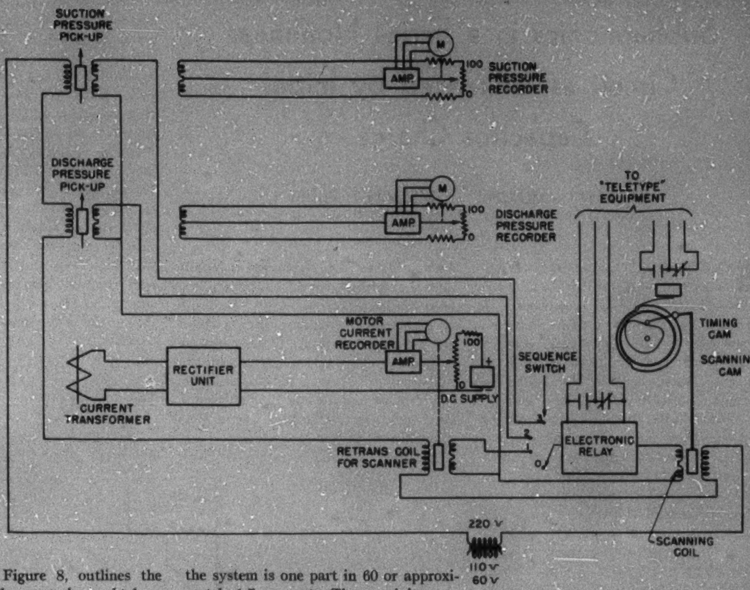
Pipeline Application

This system was first used on automatic pumping stations for a pipeline of the Shell Oil Company. It is installed, as shown in Figure 6, on the four automatic booster stations located at New Douglas, Dennison, and Effingham, Ill., and North Salem, Ind. The stations make a much higher throughput possible than usual in an 8-inch line as well as increasing overall operating efficiency.

The operation of the four stations is entirely controlled from the central dispatch center in New York City. Various products are pumped in turn through the line and it has been found that the slight intermixing produces no harmful results in the quality of any of the products. The location of these products is traced on a control board which simulates the pipeline. Thus, the dispatcher in New York has a completely integrated picture of the location and quantity of all products in the whole pipeline as well as the pressures and pump motor current of each of the four booster stations. Since the dispatcher has control of these stations, as well as complete operating information, he is in a position to operate the line in a more efficient manner.

The telemetering equipment, as previously described, furnished information regarding suction pressure, discharge pressure, and pump motor current. A block diagram, Figure 7, indicates the telemetering equipment supplied for each of the four booster stations. The suction and discharge pressures are measured by gauges installed in the pump house and locally telemetered to the control room where they are recorded. A motor current recorder also is located on the same panel. Electrically isolated voltages from the three transmitting differential transformers are connected to the sequence relay and in turn to the scanner for long distance telemetering. A simplified sche-

Figure 8. Simplified schematic diagram of telemetering system



matic diagram, Figure 8, outlines the circuits used in the measuring and telemetering equipment supplied by Bailey Meter Company.

To obtain readings of telemetered variables, the dispatcher at New York City dials a code number which operates the equipment at any one of the four chosen stations and automatically telemeters the pressures and pump motor current in the form as shown previously. At the end of one telemetering cycle of the three variables, the equipment automatically shuts off and the dispatcher, if he desires, then may obtain similar information from the three other stations. The dispatcher also may start and stop the main pump motor, lubricating oil pump, suction and discharge valves, and other necessary equipment. In an emergency, he also may sound a siren to summon the operator who is located at each station for maintenance and emergency purposes.

All the information received in the New York dispatcher's office is simultaneously printed on all the other teletype machines which are connected to the same system. The teletype machine is operated at a speed of 368 characters per minute or 163 milliseconds per character. The total cycle period for the scanning of one measurand is 15 seconds, or 45 seconds for a complete intelligence report from any one station. The 100-per-cent measurement time was selected at 60 characters or 9.783 seconds. Thus, the resolution of the system is one part in 60 or approximately 1.7 per cent. The remaining portion of the cycle, 5.2 seconds, is allowed for the various switching operations involved to transfer from one pickup to the next, change code letters, and so forth. If a lower resolution were permissible, the cycle time could be reduced proportionally, and similarly for a higher resolution, a longer cycle would be employed.

Operating Experience

This system has been in operation since September 1950. Some trouble was experienced at the beginning with servo-amplifier failure, caused by faulty components, but this was easily remedied since these units are standard plug-in assemblies. Since then the equipment has been found to be completely reliable due to its inherently simple nature, lack of delicate components, and over-all ruggedness. It serves its purpose admirably in integrating all information and supervisory control at one centralized point and thus permitting much more efficient utilization and operation of the pipeline system.

Conclusions

It will perhaps appear obvious to the electronic engineer in the field of mobile telemetering that the utilization of electronic techniques could have been made in any or all of the elements of the system described. For example, an electronic scanning system could have been used in place of the electromechanical system, resulting in much greater speed of scanning. Acceptance of complex electronic systems for the control and telemetering of industrial process variables has been somewhat slow, particularly where the requirements do not exclude the use of simpler electromechanical mechanisms. Where high speed and sensitivity are required and a great number of variables are to be telemetered, the electronic systems probably would be required.

References

1. TELEMETERING, SUPERVISORY CONTROL, AND ASSOCIATED CIRCUITS, Joint Subcommittee of the AIEE Committees on Instruments and Measurements and on Substations. *Special Publication S-12, AIEE*, September 1948.

2. JOINT AIEE—NTF CONFERENCE ON TELEMETERING. *Special Publication S-41, AIEE*, August 1950.

3. ELECTRONIC-TYPE INSTRUMENTS FOR INDUSTRIAL PROCESSES, P. S. Dickey, A. J. Hornfeck. *Transactions*, The American Society of Mechanical Engineers (New York, N. Y.), volume 67, July 1945, pages 393–98.

4. PRINCIPLES AND METHODS OF TELEMETERING (book), P. A. Borden, G. M. Thynell. Reinhold Publishing Corporation, New York, N. Y., 1948.

No Discussion

Subharmonics in a Series Nonlinear Circuit as Influenced by Initial Capacitor Charge

WILLIAM J. McKUNE
ASSOCIATE MEMBER AIEE

MELVIN F. BRUST
ASSOCIATE MEMBER AIEE

SERIES circuits containing a capacitance and nonlinear inductance have assumed considerable importance in modern engineering design. However, knowledge of the behavior of such circuits is extremely limited because of the inherent nonlinearity of the differential equations encountered.

While considerable work has been done by the investigators of nonlinear vibrations and celestial mechanics on nonlinear differential equations in which the nonlinearity is involved in the damping term (the equation of van der Pol) or in the displacement term (Duffing's equation) or in the forcing function, very little has been done on differential equations in which the nonlinearity is involved in the inertia term. The subject is further complicated by the fact that transients in such networks may be extremely long so that solutions by a differential analyzer are lengthy and subject to cumulative errors.

Hence it is felt that at present the most reliable method of predicting with a reasonable degree of accuracy the performance of proposed circuits is from a wealth of experimental data. The data now available give considerable insight into the type of response that may be encountered. From the work of Rouelle,[1] McCrumm,[2] and others,[3–5] it is known that three different types of current response may be encountered in a series circuit containing a resistance, a capacitance, and a nonlinear inductance upon shock excitation. The first of these is the normal fundamental-frequency exciting current of the inductance. The second is a much higher value of current of fundamental and higher frequencies, which has been labeled ferroresonance. The third type of response contains currents which are integer submultiples of the applied frequency. This phenomenon is called subharmonic response, and the order of this response indicates the ratio of the applied frequency to the lowest frequency component of the resulting current.

The work of these investigators gives information as to the ranges of circuit parameters for which the different types of response may be obtained. Angello[6] reported the effect of the initial switching angle on the occurrence of subharmonics with constant initial capacitor charge. However, the effect of initial capacitor charge or of initial flux linkages has not been investigated by previous authors. The purpose of the investigation reported in this paper was to determine the effect of the initial capacitor charge on the occurrence of subharmonics, with the initial switching angle and initial flux linkages held constant. The investigation was restricted to third-order subharmonics as this is the order most likely to be encountered.

Test Circuit

The circuit investigated is shown in Figure 1. Figure 1(A) shows the non-

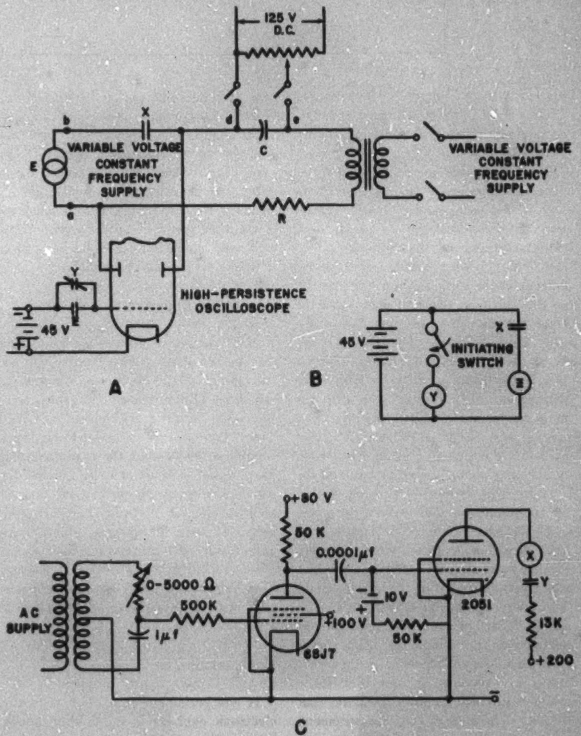


Figure 1. Test circuit

Paper 52-207, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted August 10, 1951; made available for printing May 9, 1952.

WILLIAM J. MCKUNE is an assistant professor and MELVIN F. BRUST is a graduate student at the University of Texas, Austin, Texas.

This paper is based on a thesis submitted by Mr. Brust in partial fulfillment of the requirements for a degree of Master of Science in Electrical Engineering at The University of Texas, 1951.

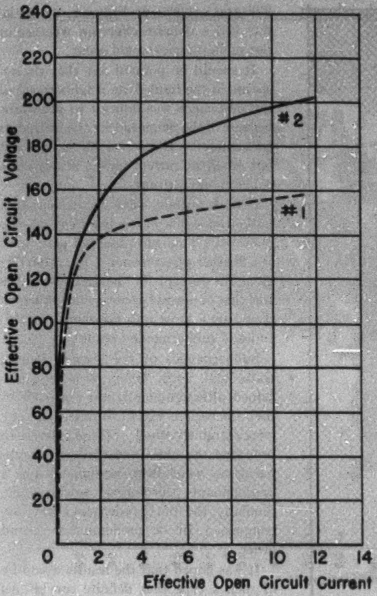


Figure 2 (left). Saturation curves of inductive elements at fundamental frequency

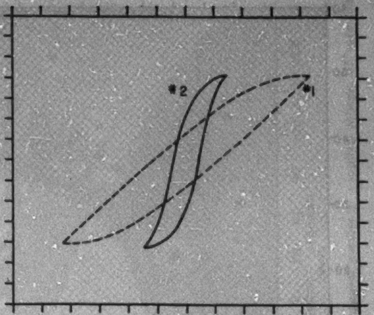


Figure 3 (right). Hysteresis loops at 57 volts rms

linear circuit containing a resistance R, a capacitance C, and a nonlinear inductance (a transformer with open secondary) connected through a relay contact X to a low-impedance variable voltage source. The capacitor was charged prior to closing the relay contact by applying a unidirectional voltage. Inasmuch as the occurrence of subharmonics was found to be influenced by initial flux linkages, the transformer was demagnetized by applying a decreasing-amplitude sinusoidal voltage to the secondary of the transformer prior to each reading.

A high-persistence oscilloscope was used to set the relay contact X to close at the desired angle and to monitor this angle continuously. To observe the angle it was necessary to blank the oscilloscope approximately one cycle after relay X operated. It was found that if the beam of the oscilloscope was left on prior to operation of the X relay the initial trace on the screen was objectionable. This was eliminated by applying a negative voltage to the grid of the cathode-ray tube, removing this bias just prior to the closing of X, and putting it back approximately 1 cycle later. Because of the high potential to ground of the grid circuit, it was necessary to isolate this circuit from the X relay.

The circuit used for control of the switching angle is shown in Figures 1(B) and 1(C). The grid of the *6SJ7* tube was supplied through a phase-shifting network such that the phase position of the approximately square wave output of the *6SJ7* tube could be varied with respect to the applied voltage. This output was applied to the grid of the thyratron through a differentiating network with a time constant of about 5 microseconds. The phase position of the resulting pulse could thus be regulated by means of the phase shifter in the grid of the pentode. The magnitude of the pulse was controlled by adjusting the input to the phase shifter and was set to such a value that the positive pulse would raise the grid voltage of the thyratron slightly above cutoff. Thus the thyratron was triggered once every cycle with a pulse of very short duration and of variable phase position and would fire at this point if the plate circuit was closed.

Method of Taking Data

The data were taken as follows. The residual flux in the inductance was destroyed and the desired potential was placed on the capacitor. As soon as the initiating switch was closed, relay Y closed, which unblanked the oscilloscope and completed the plate circuit to the thyratron. The next positive triggering impulse from the pentode fired the thyratron, energizing relay X. Relay X then energized the nonlinear circuit and picked up relay Z which again blanked the oscilloscope. The proper time interval was obtained by adjusting the spacing of the relay contacts.

Since the work of Angello[6] indicated that the probability of subharmonic response was greatest at zero switching angle, it was decided to energize the circuit at this point. The wave form of the current in the nonlinear circuit was observed on a low-persistence oscilloscope connected across the resistive element.

Two different inductances were used in order to determine the effect of different saturation curves. The saturation curves for the two transformers are shown in Figure 2. The hysteresis loops for the two transformers at an applied voltage of 57 volts rms, 60 cycles, are shown in Figure 3.

Three tests were made with transformer number 2: one with constant capacitance and variable applied voltage, and two with constant applied voltage and variable capacitance. With transformer number 1, one test was made with constant capacitance and variable voltage and one with variable capacitance and constant applied voltage. The values of voltage and capacitance which were held constant were selected because in a rough exploration of the complete regions of response they proved to be the values most conducive to subharmonic response.

The results obtained are shown in Figures 4 through 8 and Appendix I shows the conditions under which they were obtained.

Conclusions

Several facts are worthy of note. In his experimental work Rouelle[1] found

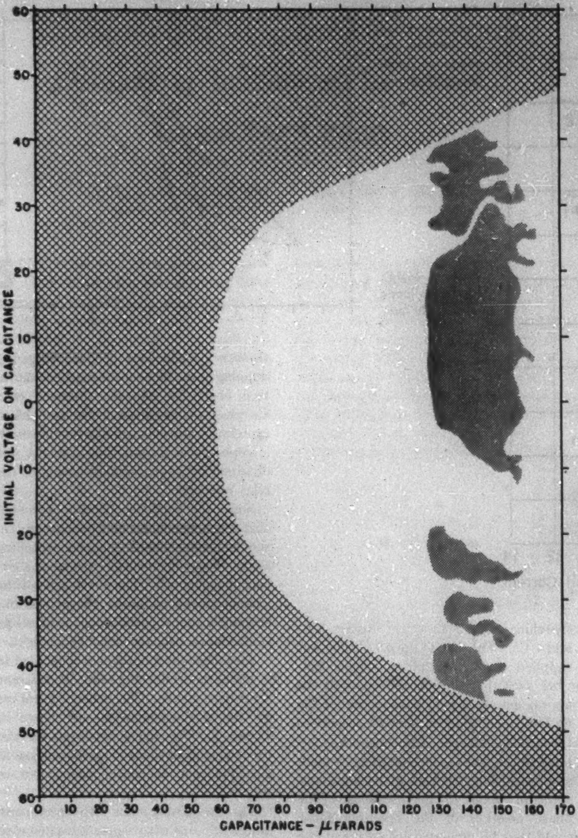


Figure 4. Areas of stable response. Transformer number 1—applied voltage, 60 volts rm

that, for random excitation, subharmonic response could not be obtained as easily for a transformer with a sharply curved saturation curve as for one in which the saturation curve had less severe curvature.

The investigation herein described supports this finding as can be seen from a comparison of the areas for the two transformers.

In his work McCrumm[2] found that for certain regions of capacitance, subharmonics could not be obtained for certain definite values of applied voltage although they could be obtained for values of applied voltage above or below this value. Dehors' work[7] supported this conclusion. However, although a lengthy search was made on the two inductances used in the present investigation, this gap could be found on neither. One possible explanation is that transformer number 2 had a high value of conductor resistance; McCrumm pointed out that high values of resistance could destroy this "lake." And the curvature of the saturation curve for transformer number 1 may have been so sharp that this phenomenon may not have existed.

Angello[6] found that the initial flux linkages in the inductances had no effect on the wave forms of stable subharmonics. This was confirmed; however, the initial flux had a definite effect on whether or not subharmonics would occur.

It should be pointed out that the response in the immediate neighborhood of ferroresonance was difficult to determine because of the magnitude of the transients encountered. In no case did a transient ferroresonant current yield a steady-state subharmonic current.

Several authors have reported that a prolonged subharmonic transient may be obtained. This was observed here, with yet another phenomenon. In certain regions the current for the first few cycles was that of normal exciting current which would then grow into the higher magnitudes of subharmonic response.

Subharmonics of the second, fourth, sixth, and ninth orders were also obtained, although no attempt was made to map the regions except to note that they were extremely small. It is of interest to note that the sixth- and ninth-order subharmonics were always accompanied by a predominant third-order subharmonic. Similarly, the fourth order was always accompanied by a predominant second order.

It was hoped that the results would be of such a type that definite conclusions could be reached regarding a means of avoiding ferroresonance and subharmonics by a proper initial charge on the capacitor. In the case of subharmonics no such conclusions could be drawn. It can be seen, however, that by putting zero charge on the capacitor, ferroresonance is not likely to be obtained whereas it may be obtained with a high value of initial charge. The placing of a resistor in parallel with the capacitor as suggested by Butler and Concordia[8] accomplishes this by allowing a path over which the residual charge may be dissipated.

Appendix I

Frequency of applied voltage = 60 cycles per second
Series resistance R = 0.80 ohm
Switching angle = 0 degree
Initial flux linkages = 0.0
Transformer number 1 = 2-kva distribution transformer, Hypersil core, 0.22-ohm conductor resistance
Transformer number 2 = power supply plate transformer, 2.20-ohm conductor resistance
Sign convention = switch closed with the voltage rise from a to b, see Figure 1(A), passing through zero from negative to positive
= charge on capacitor C arbitrarily defined as positive when d was positive with respect to e

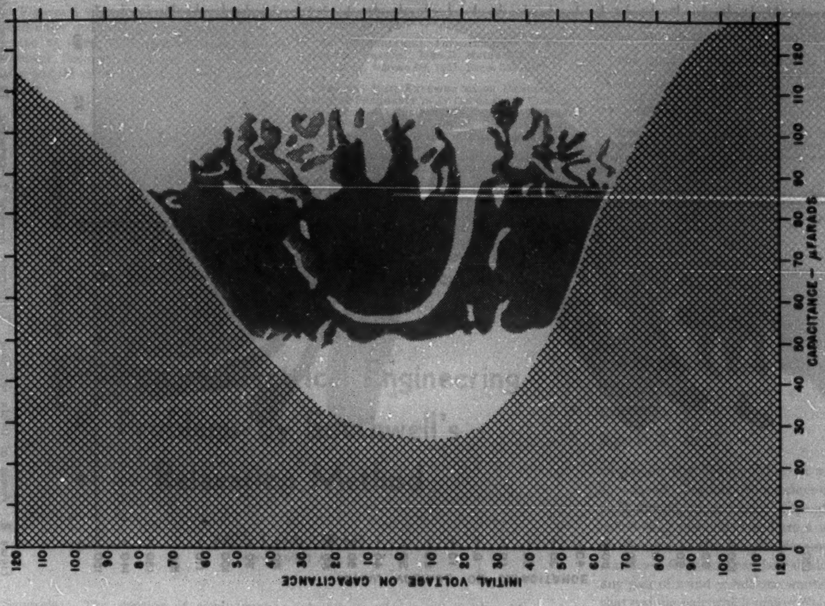


Figure 6. Areas of stable response. Transformer number 2—applied voltage, 55 volts rms

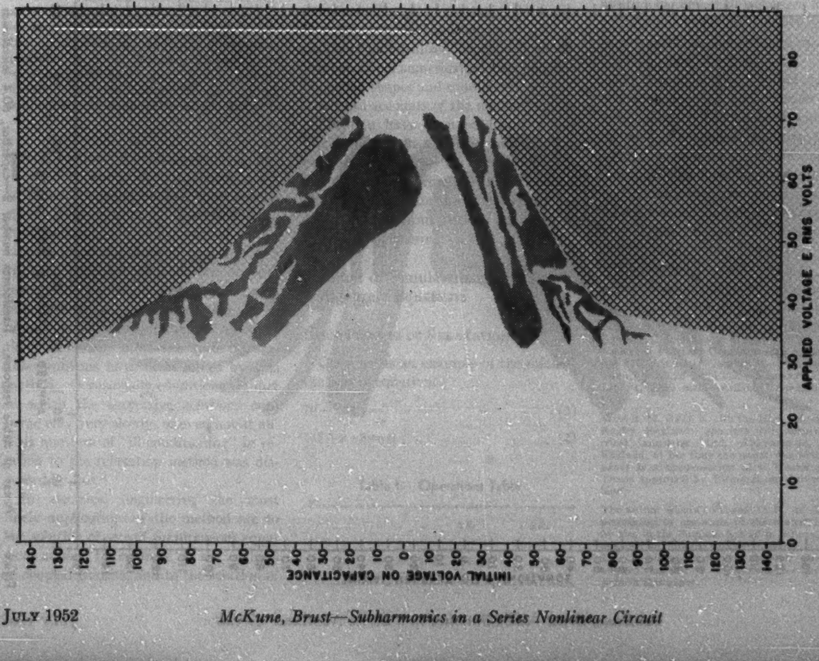


Figure 5. Areas of stable response. Transformer number 1—capacitance, 136.9 microfarads

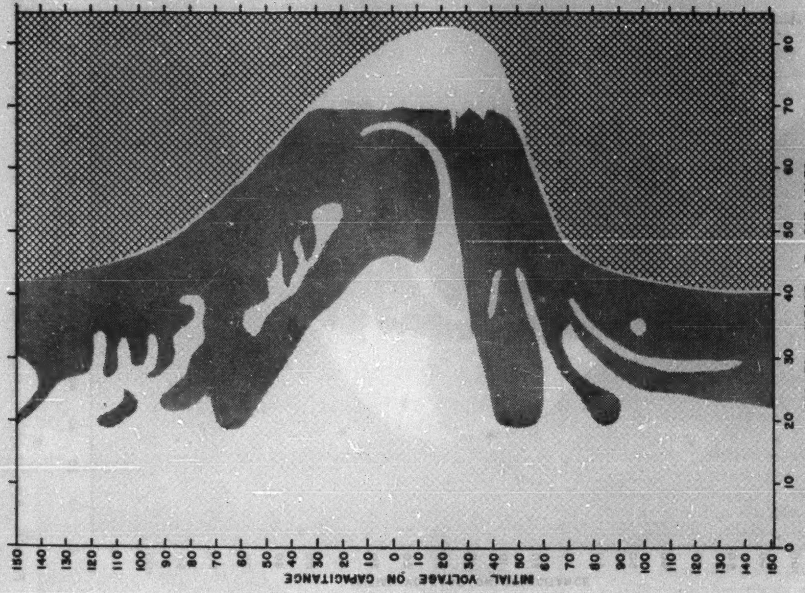


Figure 8. Areas of stable response. Transformer number 2—capacitor, 79.2 microfarads

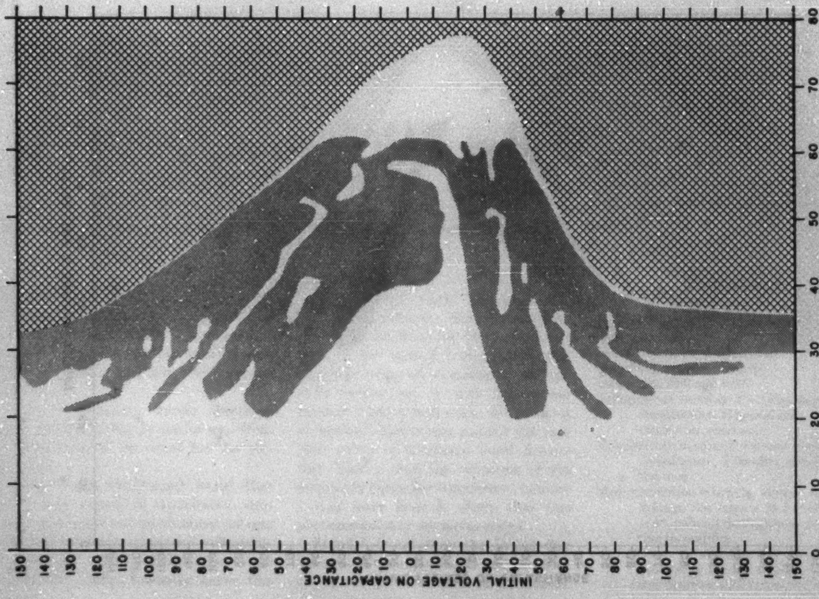


Figure 7. Areas of stable response. Transformer number 2—capacitor, 60.4 microfarads

In Figures 4 through 8, the area of stable third-order subharmonic response is represented by dots. The area of stable ferroresonant response is shown by crisscross lines and the area of normal exciting current is clear.

References

1. Contribution to the Experimental Study of Ferroresonance, E. Rouelle. *Revue Générale de l'Électricité* (Paris, France), volume 36, November 24, 1934, pages 715–38; volume 36, December 1, 1934, pages 763–80; volume 36, December 8, 1934, pages 795–819; volume 36, December 15, 1934, pages 841–58.

2. An Experimental Investigation of Subharmonic Currents, John D. McCrumm. *AIEE Transactions*, volume 60, 1951, pages 533–540.

3. Several New Experiences of Demultiplication of Frequency in an Oscillatory Circuit with a Coil Having an Iron Core, E. Rouelle. *Revue Générale de l'Électricité* (Paris, France), volume 40, December 26, 1936, pages 811–819.

4. On a Static Demultiplier of Frequency, J. Fallou. *Revue Générale de l'Électricité* (Paris, France), volume 19, June 19, 1926, pages 987–991.

5. On Certain Peculiarities of Ferroresonance When Resistance Is Not Negligible, E. Rouelle. *Comptes Rendus Hebdomadaires des Sciences de l'Academie des Sciences* (Paris, France), volume 184, June 13, 1927, pages 1426–28.

6. The Effect of Initial Conditions on Subharmonic Currents in a Nonlinear Series Circuit, Stephen J. Angello. *Electrical Engineering* (*AIEE Transactions*), volume 61, September 1942, pages 625–27.

7. Research on the Ferromagnetic Demultiplication of Frequency, R. Dehors. *Revue Générale de l'Électricité* (Paris, France), volume 56, November 1947, pages 455–67.

8. Analysis of Series Capacitor Application Problems, J. W. Butler, C. Concordia. *Electrical Engineering* (*AIEE Transactions*), volume 56, August 1937, pages 975–88.

No Discussion

Solution of Electrical Engineering Problems by Southwell's Relaxation Method

MISS E. M. GRAD
NONMEMBER AIEE

Synopsis: The solution of simultaneous linear algebraic equations by the basic relaxation method is described in detail and the use of short-cut methods is explained and demonstrated. It then is shown that the same technique can be applied to the solution of ordinary and partial differential equations, and its accuracy is discussed. Several problems taken from various branches of electrical engineering are solved to illustrate the application of the method.

THE method of "relaxation" as devised by Southwell[1–8] for the solution of framework problems has proved to be of very wide application. Used first in the numerical solution of systems of simultaneous linear algebraic equations, it now has been extended, with outstanding success, to the solution of ordinary and partial differential equations. The technique is essentially the same for all equations, and many otherwise quite intractable problems have been solved by this method. Occasionally conditions are met in which the successive solutions converge only very slowly, or even not at all. This question of "ill-conditioning" in relation to the relaxation method was discussed by Fox.[4]

In electrical engineering the most likely applications of the method are to the solution of sets of simultaneous equations arising, for example, in the analysis of coupled circuits, and to the solution of electrical and magnetic field problems, determined by the partial differential equations of Poisson and Laplace, or the wave equation. By other methods, these differential equations can be solved only in simple or in "idealized" cases, but relaxation can give a numerical solution for most boundary shapes and conditions.

Several accounts of the method and its application have been published[1–7] but few deal with the actual technique. The purpose of this paper is to present an outline of the basic principles of the method and to demonstrate the technique by examples taken from various branches of electrical engineering.

Solution of Simulataneous Linear Algebraic Equations

Basic Process of Relaxation

Consider as an example of the method the pair of equations

$$79-3x+y=0 \tag{1}$$

$$147+x-3y=0 \tag{2}$$

Table I. Operations Table

	ΔR_1	ΔR_2
$\Delta x=1$	-3	1
$\Delta y=1$	1	-3

These equations may be written in the form

$$R_1\equiv 79-3x+y \tag{3}$$

$$R_2\equiv 147+x-3y \tag{4}$$

where, for the correct values of x and y, $R_1=R_2=0$. For any other values of x and y, R_1 and R_2 are not zero. R_1 and R_2 are called the "residuals" and are a measure of the error of the solution for any given x and y. R_1 and R_2 can be calculated for any pair of x and y, but one wants to find that unique pair which makes $R_1=R_2=0$.

First, an operations table is set up. From equations 3 and 4 it is seen that unit positive increment to x changes R_1 by -3 and R_2 by $+1$ and unit positive increment to y changes R_1 by $+1$ and R_2 by -3. This is given in Table I.

One has to make an initial guess of x and y to start the computation (a good initial guess is helpful as it will shorten the subsequent numerical work, but is by no means essential). In this particular example, take for convenience $x=0$ and $y=0$ as initial values. From this one calculates the residuals R_1 and R_2 using equations 3 and 4 and enters these values in the relaxation table, Table II.

The biggest residual then is picked out, which is 147 in this case, and it reduces to

---

Paper 52-213, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted July 24, 1951; made available for printing May 15, 1952.

Miss E. M. Grad is with the High School for Girls, Rugby, England; she was with Associated Electrical Industries., Ltd., Aldermaston, Berkshire, England, at the time the paper was written. This paper is a condensation of a Master of Science Thesis approved by Birmingham University, England.

The author wishes to thank D. N. de G. Allen for permission to use some of the material presented by him in the Summer School in Relaxation Methods held at Imperial College, London, in June 1948; Dr. G. Liebmann for his constant guidance and advice; and Dr. T. E. Allibone for permission to publish this paper.

Table II. Relaxation Table

		R_1	R_2
$x=0$	$y=0$	79	147
	$\Delta y=49$	128	0
$\Delta x=42$		2	42
	$\Delta y=14$	16	0
$\Delta x=5$		1	5
	$\Delta y=1$	2	2
$\Delta x=1$		−1	3
	$\Delta y=1$	0	0
$x=48$	$y=65$		

Check $R_1=79-144+65=0$.
$R_2=147+48-195=0$.

zero. Since $\Delta y=1$ gives $\Delta R_2=-3$, $\Delta y=49$ gives $\Delta R_2=-147$, thereby making R_2 zero, but at the same time it makes $\Delta R_1=49$, so increasing R_1 to 128. This operation is entered in the second line of the Relaxation Table. Proceeding in the same way, take $\Delta x=42$, giving $\Delta R_1=-126$. This does not make R_1 quite zero, but as near to zero as necessary. The corresponding increment to $R_2=42$, and this operation is written as line 3. The working is continued in this manner until both residuals are zero. (In line 7 a negative residual occurs, but this presents no difficulty.) The increments to x and y then are summed and the values so obtained substituted in the original equations, to check that $R_1=R_2=0$.

If a mistake had been made in Table II, R_1 and R_2 would not be found to be zero, but some other small numbers. These values then would be taken as new starting values for the residuals and "relaxed" to zero, the further increments required to do this being added to the previous ones to give the correct solution.

Relaxation is used to get a numerical solution as quickly as possible. The method involves a large number of very simple operations. Since the results can be checked easily during the progress of the work, it is better to work quickly and risk arithmetical errors, making checks of the residuals only occasionally, than to stop and check each step by which the residuals have been calculated. To make the method even more speedy, several shortening devices have been developed.

SHORTENING DEVICES

Overrelaxation. From the example it can be seen that whenever one residual is decreased the other is increased though by a smaller amount. If, therefore, instead of making just sufficient increment to make a residual zero, one makes a larger increment, that is, one overrelaxes to make the residual negative, then when it receives an increment as a result of another operation it still will be near zero. This can be seen from the alternative method, Table III, of working the previous example.

Table III. Alternative Method

		R_1	R_2
$x=0$	$y=0$	79	147
	$\Delta y=60$	139	−33
$\Delta x=45$		4	12
	$\Delta y=5$	9	−3
$\Delta x=3$		0	0
$x=48$	$y=65$		

Table III, of course, gives the same answer but it took fewer lines of working.

Block Relaxation. By definition this is the application of unit increment to more than one of the variables at a time. That is a superposition of the effects $\Delta x=1$ and $\Delta y=1$, giving a new line in the operations table shown in Table IV.

Table IV

	ΔR_1	ΔR_2
$\Delta x=1$	−3	1
$\Delta y=1$	1	−3
$\Delta x=\Delta y=1$	−2	−2

When the total residual is zero, individual residuals must be evenly distributed among positive and negative residuals, and such residuals are much easier to "liquidate" than a set which all have the same sign. Block relaxation is very useful in reducing the total residual (that is, the sum of all residuals) to zero. To demonstrate this the same example is again solved.

Table V

		R_1	R_2
$x=0$	$y=0$	79	147
$\Delta x=56$	$\Delta y=56$	−33	35
	$\Delta y=9$	−24	8
$\Delta x=-8$		0	0

Group Relaxation. By definition this is the application of unequal increments to more than one of the variables at a time.

Any new line can be formed in the operations table to meet a particular requirement; an example is given in Table VI.

This particular operation would be used if one residual were very large compared with the other.

Equalisation of Residuals. In any set of linear equations, if the initial residuals are $R_0(x)$, $R_0(y)$, $R_0(z)$ . . . it is possible by applying appropriate increments $\Delta_n x$, $\Delta_n y$, $\Delta_n z$ . . . to reduce these values to the values $R_n(x)$, $R_n(y)$, $R_n(z)$ . . . such that

$$\frac{R_n(x)}{R_0(x)}=\frac{R_n(y)}{R_0(y)}=\ldots=K(0<K<1)$$

It is then found that further increments $(K/1-K)\Delta_n x$, $(K/1-K)\Delta_n y$, $(K/1-K)\Delta_n z$ . . . will reduce all the residuals to zero.

APPLICATION OF THE METHOD TO COUPLED CIRCUITS

For the general case of a p-mesh linear network, the steady state is described by the p equations

$$E_1=Z_{11}I_1+Z_{12}I_2+\ldots+Z_{1p}I_p$$
$$E_2=Z_{21}I_1+Z_{22}I_2+\ldots+Z_{2p}I_p$$
$$\cdot$$
$$\cdot$$
$$\cdot$$
$$E_p=Z_{p1}I_1+Z_{p2}I_2+\ldots+Z_{pp}I_p$$

where

E's are impressed voltages (constant)
Z's are impedances (constant)
I's are mesh currents (unknown)

The currents then can be found by solving this system of simultaneous linear algebraic equations. While the solution can be obtained by elimination, it is found that in most cases relaxation is simpler and quicker to carry out.

Relaxation has the added advantages that

1. It very quickly gives an approximate solution which may be all that is required.
2. Starting from this approximate solution, a solution of any desired accuracy can be obtained.
3. The progress of the solution can be checked easily by recalculating the residuals from the original equations.

Its disadvantage is that all the currents must be determined simultaneously whether they are all of interest or not.

APPLICATION OF THE METHOD TO CHARACTERISTIC VALUE PROBLEMS

Systems of simultaneous equations dealing with characteristic value problems, for example, vibrating engineering structures, can be solved by a combination of the relaxation process as described, and of Rayleigh's principle. This method has been discussed fully in a paper by Fox[4]

Table VI

		ΔR_1	ΔR_2
	$\Delta x=1$	−3	1
	$\Delta y=1$	1	−3
$\Delta x=2$	$\Delta y=1$	−5	−1

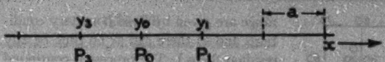


Figure 1. Subdivision of range into finite intervals

in which examples are given showing how certain devices can be used to shorten the process, and also how modes other than the lowest can be found.

Solution of Ordinary Differential Equations

NATURE OF THE PROBLEM

Ordinary differential equations of the second order can be written in the form

$$\frac{d^2y}{dx^2}=z\left(x,y,\frac{dy}{dx}\right)$$

where z is a known function of x, y, and dy/dx. For a unique solution of the equation two boundary conditions also must be specified. Sometimes the forms of the equation and boundary conditions are such that an analytical solution is impossible and then numerical methods must be used. Usually, either y and dy/dx are known at one end of the range of solution or y is known at each end. In the first of these cases, an initial-value problem, the problem is usually solved by step-by-step methods, while in the second case, a boundary-value problem, relaxation is a much better method.

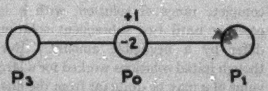


Figure 2. Incomplete relaxation pattern (first step)

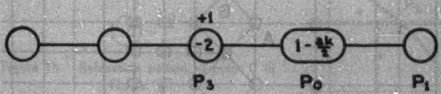


Figure 3. Incomplete relaxation pattern (second step)

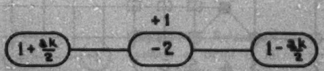


Figure 4. Complete relaxation pattern

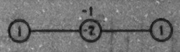


Figure 5. Relaxation pattern for differential equation $y''-z(x)=0$

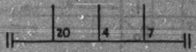


Figure 6. Assumed initial values of residuals

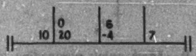


Figure 7. Increments and resulting residuals after first relaxation step

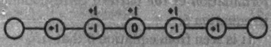


Figure 8. Relation pattern for block relaxation

Although differential equations require for their solution a continuous function in the given range, relaxation, like all other numerical methods, gives only numerical values at a series of points. These points are taken for convenience as equidistant points along the x-axis, and values at intermediate points then can be found by graphical or algebraic interpolation.

Relaxation cannot solve differential equations as such, but by replacing them by the corresponding finite-difference equations they can be converted into systems of simultaneous linear algebraic equations which are amenable to the standard relaxation techniques, provided that the solution does not contain singular points within the range of the solution.

Differential equations leading to characteristic values and characteristic solutions can be dealt with as mentioned previously in the section entitled "Application of the method to characteristic value problems" after the differential equation has been replaced by a set of simultaneous linear equations.

FINITE DIFFERENCE APPROXIMATION

Consider, for example, the equation

$$\frac{d^2y}{dx^2}+K\frac{dy}{dx}-z(x)=0$$

with the boundary conditions that y is known at each end of the range of x under consideration. This equation, with K being a constant, of course can be solved directly and the appropriate boundary conditions inserted, but it was chosen here as an example to demonstrate the method in a simple way. One divides this range of x into equal mesh lengths a. Taking any nodal point as the origin $P_0(x_0)$ where y has the value y_0, and adjacent points $P_1(x_0+a)$ and $P_3(x_0-a)$ where $y=y_1$ and $y=y_3$ as shown in Figure 1, one can show from Taylor's expansion that the finite difference approximations are

$$\left(\frac{dy}{dx}\right)_0=\frac{y_1-y_3}{2a}$$

and

$$\left(\frac{d^2y}{dx^2}\right)_0=\frac{y_1+y_3-2y_0}{a^2}$$

The given equation therefore can be approximated by

$$y_1(2+ak)+y_3(2-ak)-4y_0-2a^2z_0=0$$

This is a linear algebraic equation, and since an equation of this form connects every set of three consecutive points, the given differential equation has been replaced by a set of simultaneous linear algebraic equations, the number of equations in the set being given by the number of points at which the solution is determined.

The residual R_0 at the point P_0 is given by

$$\boxed{R_0=y_1(1+ak/2)+y_3(1-ak/2)-2y_0-a^2z_0} \qquad (5)$$

BASIC PROCESS OF RELAXATION

Since all the equations 5 (for different points $P_0, P_1 \ldots P_n$) are of the same form, every line of the operations table is of the same form. It is best visualized not as a table but as a "relaxation pattern."

This relaxation pattern is a graphical representation of the effect on residuals at various points of unit increment at any given point. Patterns can be formulated for any residual formula, and they are the essential feature of the relaxation method. Once the pattern has been made, it can be applied to the numerical work in which all the residuals are reduced to negligible values. It is left to the discretion of the computer to decide at which points these patterns will be used most profitably. This makes relaxation a more interesting and more flexible computational technique than the purely mechanical iteration methods.

The character of the relaxation pattern depends on the form of the given differential equation. It may be the same for all points in the region, or it may depend on the position of the origin. The method of deriving the pattern, however, is the same

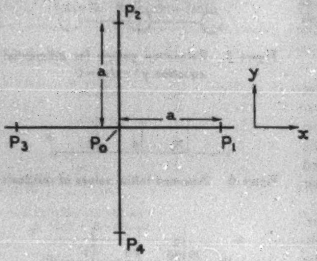


Figure 9. Mesh points surrounding point P_0

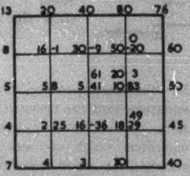


Figure 12. Potential values, increments, and residuals after first relaxation step

in all cases. As an example, consider the formulation of the pattern corresponding to equation 5.

It is seen from equation 5 that $\Delta R_0 = -2$ for $\Delta y_0 = +1$. This is represented graphically by the (incomplete) pattern, Figure 2, the increment Δy being written above the circle, the resulting change of residual inside the circle. However, the increment Δy_0 does not only change the residual at the point P_0 but also the residuals R_1 and R_3 at the neighboring points P_1 and P_3. The changes of the residuals at these points due to the increment $\Delta y_0 = 1$ are obtained from equation 5 by changing the origin first to P_1 and then to P_3. Taking the new origin at P_1, the value y_0 takes the place of y_3 in equation 5, and hence the increment Δy_0 produces in the residual R_1 the change $\Delta R_1 = (1 - ak/2)\Delta y_0$. This is shown graphically in Figure 3 for a unit increment $\Delta y_0 = 1$. Similarly one finds that the change in the residual R_3 is $\Delta R_3 = (1 + ak/2)\Delta y_0$. The resulting complete relaxation pattern, representing the change of all residuals due to a unit increment of y at P_0, is shown in Figure 4.

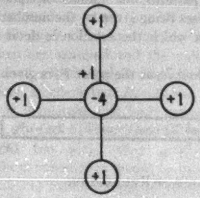


Figure 10 (left). Relaxation pattern for Laplace's equation

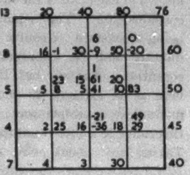


Figure 13 (right). Potential values, increments, and residuals after further relaxation

Consider now the simple example where $K = 0$. Here the relaxation pattern reduces to that shown in Figure 5. Suppose the residuals are initially as shown in Figure 6. As in the case of simultaneous equations, one starts by picking out the largest residual 20. From the relaxation pattern one sees that to make this zero one has to make an increment $+10$, so changing this residual by $10x(-2)$ and those on either side of it by $10x(+1)$. This operation is shown in Figure 7, the increments being written to the left of the vertical lines, the new values of the residual to the right. (The ends of the range are given by fixed boundary conditions and so there can at no time be any residual there.) The process is continued until all the residuals may be deemed negligible, for the accuracy desired. It is also necessary to impose the condition that the total residual shall be small.

Shortening Devices

As before, overrelaxation is quite straightforward and very useful.

Patterns for block relaxation are formed by superposing individual patterns and are seen to be of the form shown in Figure 8.

Patterns for group relaxation can also be formed to suit any particular case; see the section on "Numerical Example" given later in the paper.

Accuracy of the Method

The difference equations can be solved to any degree of accuracy but the ultimate accuracy of the solution depends on the closeness of the approximation of these difference equations to the given differential equation. Since terms of order a^3 were neglected in this approximation, the solution can be accurate only when worked for values of a which are small enough for these terms to be truly negligible. The usual practice is to work solutions for $a = (x_2 - x_1)/2n (n = 1,2,3 \ldots)$ where a is the mesh length and $x_2 - x_1$ the complete range of solution, with n increasing until two subsequent solutions differ by a negligible amount. Even though initial solutions worked for a large value of a may be quite far from the final, correct solution, they are useful and time-saving in giving a good initial guess when the interval is halved, and the convergence of the solution can be seen.

As an example, the equation

$$\frac{d^2y}{dx^2} + x(1-x)10^4 = 0$$

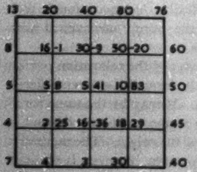


Figure 11 (left). Initial potential values and residuals for solution of Laplace's equation

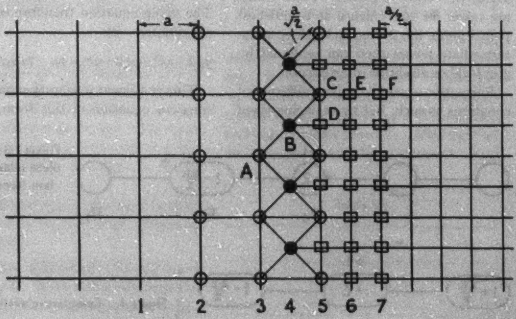


Figure 14 (right). Subdivision of relaxation nets

Table VII

n	1	2	3	4	Extra-polated Value	Analyt-ical Solution
y	313	273	264	261	260	260.4

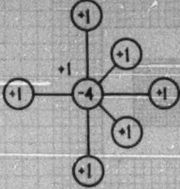


Figure 15. Relaxation pattern at point A in Figure 14

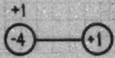


Figure 16. Relaxation pattern at point B in Figure 14

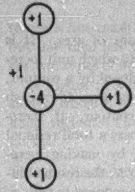


Figure 17. Relaxation pattern at point E in Figure 14

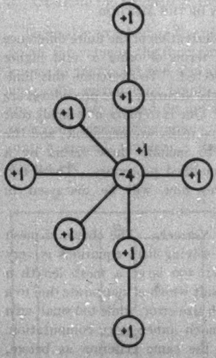


Figure 18. Relaxation pattern at point C in Figure 14

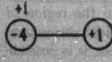


Figure 19. Relaxation pattern at point D in Figure 14

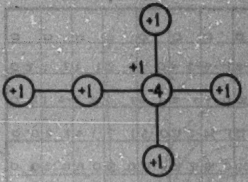


Figure 20 (left). Relaxation pattern at point F in Figure 14

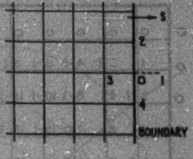


Figure 22 (right). Fictitious boundary point to represent prescribed boundary gradient

with $y=0$ at $x=0$ and $y=0$ at $x=1$, which has the analytical solution $y=260.4$ at $x=1/2$, has been solved for various n and the results are shown in Table VII.

Often, the solution approaches upon repeated reduction of the interval size a the correct value in an oscillatory manner and extrapolation must be carried out with care as was pointed out by Salvadori.[5]

Solution of Poisson's and Laplace's Equations in Two Dimensions

NATURE OF THE PROBLEM

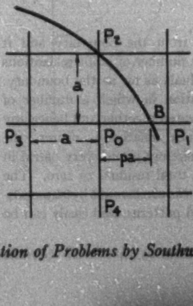
Partial differential equations may also be of initial-value or boundary-value types. As in the case of ordinary differential equations it is the boundary-value types which are solved most conveniently numerically by relaxation. Such problems arise frequently in electrical engineering. For instance, in electric machines, the shape of the boundary and the field values on the boundary are known, and the field configuration inside the region is required. Here, electric and magnetic potentials satisfy Poisson's or Laplace's equations.

The first step in solving these equations is to replace them by their finite difference approximations, that is, a set of simultaneous linear algebraic equations holding at various points at which the solution is to be obtained. The most convenient points to take are the nodes of a symmetric and systematic network; in practice, square meshes generally are used, though triangular and hexagonal meshes are geometrically possible.

FINITE DIFFERENCE APPROXIMATION

Poisson's equation

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}=s(xy)$$



may be written in finite difference form as

$$\sum_{n=1}^{4} V_n-4V_0=a^2s_0$$

a being the mesh size (see Figure 9) whence the residual R_0 at the point P_0 is given by

$$R_0=\sum_{n=1}^{4} V_n-4V_0-a^2s_0 \qquad (6)$$

For Laplace's equation $s=0$, and the equation reduces to

$$R_0=\sum_{n=1}^{4} V_n-4V_0 \qquad (7)$$

BASIC PROCESS OF RELAXATION

By arguments similar to those in the preceding section of the same name, it is found that the appropriate relaxation pattern in this case is that shown in Figure 10.

In solving Laplace's equation, for given boundary values, in the region shown in Figure 11, V values first are guessed and entered to the left of the nodes, the calculated residuals being entered to the right.

The largest residual then is relaxed as shown in Figure 12, the process continued as in Figure 13 and further continued until

1. All the residuals are small.
2. The total residual is small.
3. The total residual in any area is small.

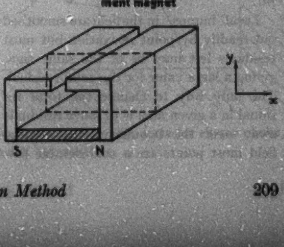
Finally, all increments ΔV are summed for each node.

SHORTENING DEVICES

Any device which will increase the speed of the method is of great importance

Figure 21 (left). Unequal network star for curved boundary

Figure 23 (below). Section through instrument magnet



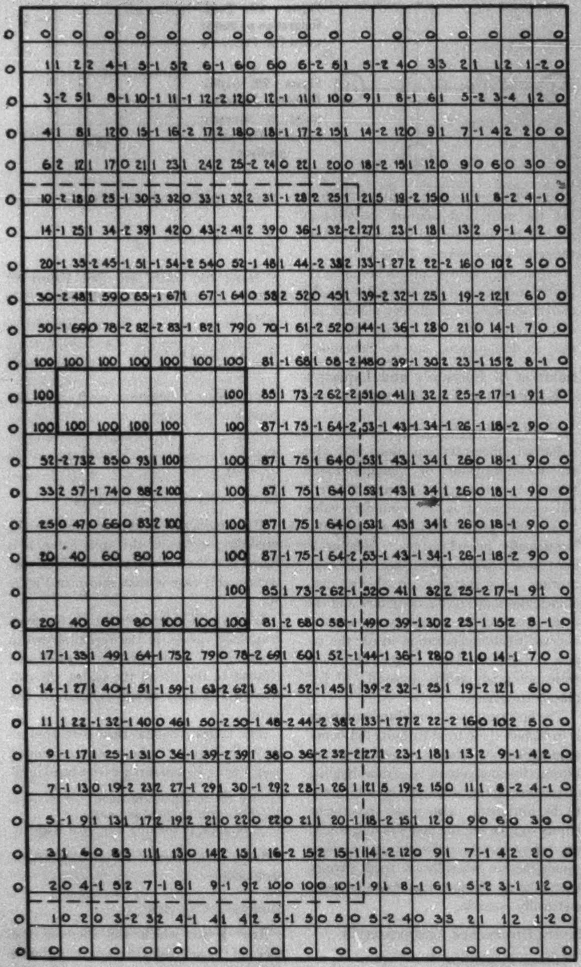


Figure 24. Relaxation net for magnet, Figure 23, showing computed field values and residuals

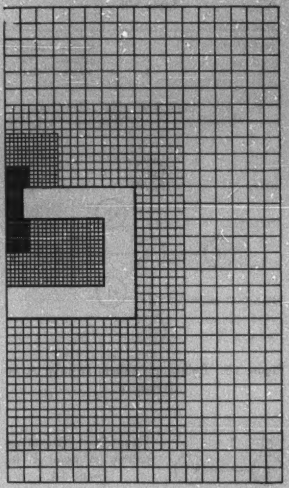


Figure 25. Refinement of relaxation net for magnet, Figure 23

and again overrelaxation proves very useful.

Local "bumps" in the field are smoothed out readily by point relaxation but most residuals left may be of the same sign, giving a large value for the total residual. The only way to change the total residual in a given area is to move residuals away over the boundary. In a large field most points are a considerable distance away from the boundary and it takes a great number of point relaxations to shift residuals as far as the boundary. Block relaxation in which a number of points are treated together and hence any point becomes relatively nearer the boundary, therefore proves very useful in reducing the total residual to zero. The patterns for this are formed by superposing individual patterns, and easily can be made to suit any shape of area. It is found that if the area to which unit increments are made is enclosed by a contour, then one unit of residual goes down each mesh line cut by the contour. If, therefore, the contour encloses a total residual R, and cuts N lines, by making increments of magnitude R/N, the residual inside the contour will be reduced to zero.

ACCURACY OF THE METHOD

In the derivation of the finite difference formulae, terms of order a^4 and higher were neglected. To overcome this limitation to the accuracy, two procedures are possible. One is to make a so small that these terms really are negligible, and the other is to include these terms by a slightly modified procedure. These two alternatives now will be discussed in turn.

Graded Networks. The choice of mesh length in solving field equations is very important; too large a mesh length a gives a result which is inaccurate due to a finite mesh size error, while too small an a requires much unnecessary computation. Adopting the same criterion as before, that is, that the solution does not change appreciably on reducing the mesh size, it frequently is found that this condition is satisfied in some parts of the field but not in others, for example, in the region of sharp corners where the field varies very rapidly. To minimize labor, it is possible

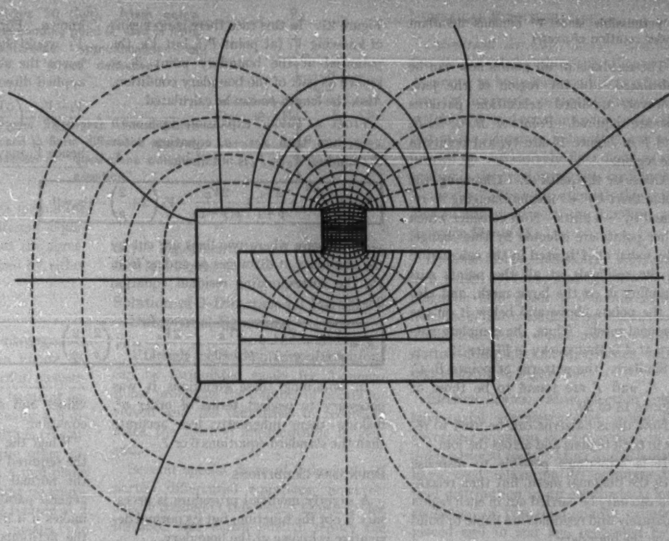


Figure 26. Magnetic flux lines and equipotentials of magnetomotive force for magnet, Figure 23, as finally computed

to use a fine mesh only where necessary, but one then has the problem of joining the two types of net.

In Figure 14 a net of mesh of size a is shown, which is to be halved in the right-hand region. The procedure is as follows:

1. Draw in the extra mesh lines required.
2. Guess values at the new nodes so formed.
3. Calculate the residuals at these points.

To be able to calculate the residual at a point, the value of the function has to be known at that point and also at the symmetrically surrounding points. Some of the nodes just formed do not have symmetrically surrounding points and so present a special problem. Referring to Figure 14:

Columns 1, 2, 3, and to the left—these are all in the coarse mesh.

Columns 6, 7, and to the right—these are all in the fine mesh.

Column 5—alternate points can be regarded as being in the coarse mesh, but the others have no point to their left. Produce the fine net one step further at these places to get the points shown in column 4.

Column 4—these points are now surrounded by points distance $a/\sqrt{2}$ away. The residuals are now calculated from equations 6 or 7 but with a replaced by $a/\sqrt{2}$. (This

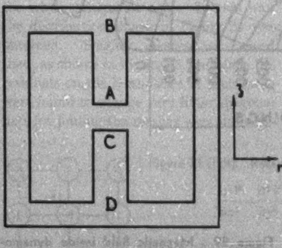


Figure 27. Section through cylindrical air-gap choke

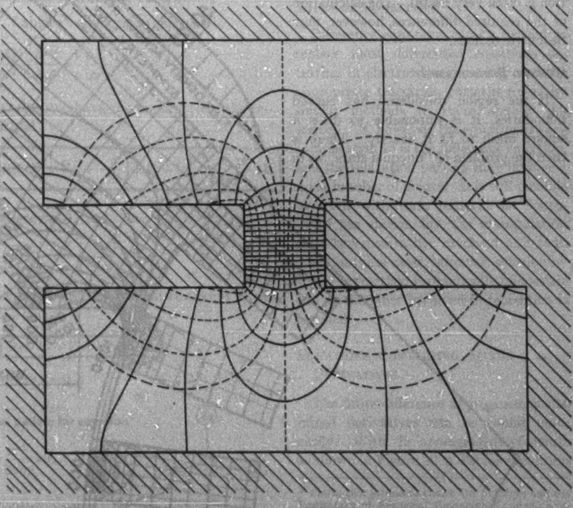


Figure 28. Magnetic flux lines and equipotentials of magnetomotive force for air-gap choke, Figure 27, as computed by relaxation

is permissible since ∇^2 remains invariant under rotation of axes.)

The residuals at all points now can be calculated. In the region of the join, however, modified relaxation patterns also are required. Points A, B, C, D, E, and F in Figure 14 are typical points in the region.

Consider the point A. Changing the value there by +1 means changing its residual by −4 units. Now consider which other points are affected by this change. The value at A is used in the calculation of the residuals at all the points surrounding it on the large mesh, and also at the points above and below it on the diagonal mesh. Thus, the complete pattern at A is that shown in Figure 15.

Similarly, the patterns at points B, C, D, E, and F are found to be those of Figures 16 to 20.

Now these patterns can be used to relax in both regions and across the join.

It is sometimes possible to dispense with the diagonal mesh, but then relaxation can only be carried out in each region separately and residuals are liable to build up on the boundary.

Inclusion of Higher Order Terms. Fox[9] has shown that by taking account of higher order terms in the finite difference approximation greater accuracy can be obtained with a relatively large value of the mesh length a.

Since the work involved increases so rapidly with a, this method of computation is often very useful. The calculation of the required differences, however, may be a little tedious.

CURVED BOUNDARIES

If the region considered has curved boundaries, it is impossible to draw a square mesh to fit it exactly. There is then the problem of "unequal stars" as in Figure 21. In this case there is no means of knowing V_1 (at point P_1), but V_B, the potential at the boundary point B, is known as part of the boundary condition. Also, the length pa can be calculated.

From a Taylor expansion first with $x=a$, and then $x=-a$, equation 8 for the residual R_0 at P_0 is obtained.

$$\boxed{R_0 = V_2 + V_4 + \frac{2V_3}{p+1} + \frac{2V_B}{p(p+1)} - V_0\left(2+\frac{2}{p}\right) - a^2 z_0 = 0} \qquad (8)$$

In the case where two lines are cut by the boundary at distances pa and qa from P_0, the corresponding residual equation 9 is

$$\boxed{R_0 = \frac{2V_3}{p+1} + \frac{2V_4}{q+1} + \frac{2V_B}{p(p+1)} + \frac{2V_c}{q(q+1)} - V_0\left(\frac{2}{p}+\frac{2}{q}\right) - a^2 z_0 = 0} \qquad (9)$$

In obtaining these equations it was necessary to neglect terms of order a^3, making them inherently less accurate than the standard equations 6 or 7.

BOUNDARY CONDITIONS

A slightly modified procedure is necessary if not the function, but its normal derivative is known at the boundary.

Consider first a rectilinear boundary as shown in Figure 22, with S the direction normal to the boundary, $\partial V/\partial S$ being known. For all points on the boundary, V_1 would have a fictitious value, and hence the standard equation cannot be applied directly. However, one may say

$$\frac{V_1 - V_3}{2a} = \left(\frac{\partial V}{\partial S}\right)_0$$

hence V_1 may be found in terms of known values and substituted in the standard equation.

When the boundary is curved, two of the required values may be fictitious, and the normal to the boundary is not in general parallel to the mesh lines. This makes it a little more difficult to express the fictitious values in terms of known ones, but it is still quite straightforward. This method of attack is due to Fox.[10] An alternative approach due to Southwell

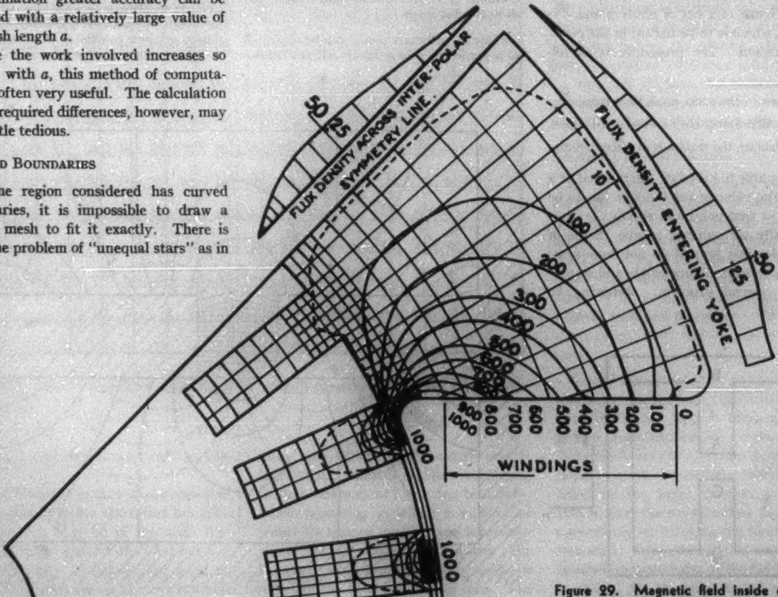


Figure 29. Magnetic field inside dynamo-electric machine computed by relaxation (Motz and Worthy)

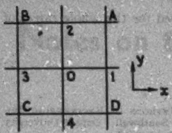


Figure 30 (left). Mesh points surrounding point 0, used in finite difference approximation of "mixed" derivatives

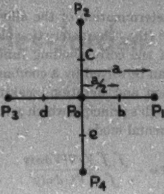


Figure 32 (right). Mesh points used in finite difference approximation of "quasi-plane potential equation"

and Vaisey[11] is based on the membrane analogy. This leads to different equations, but Fox has shown that the agreement between solutions worked by either method is quite good.

NUMERICAL EXAMPLE

Consider as an example the bar magnet with pole pieces as shown in Figure 23. This is similar to the type used in measuring instruments, and so is of frequent occurrence. The problem is really a 3-dimensional one, but if the width is reasonably large compared with the other dimensions, a section through a mid-plane can be regarded as experiencing no transverse end-effects, and can be treated as a 2-dimensional case. This means the governing equation has the form

$$\frac{\partial^2\Psi}{\partial x^2}+\frac{\partial^2\Psi}{\partial y^2}=0$$

where $H=-\text{grad}\ \Psi$, being the magnetic potential and H the field strength. The shape of the boundary is specified accurately, but the values of $H=-\text{grad}\ \Psi$ on the boundary are not known exactly. As a first approximation it is assumed that there is a linear drop of magnetic potential Ψ along the magnet and that the iron has infinite permeability.

Working on a net of mesh length one-quarter the gap width, the residuals were found to be predominantly negative giving a large value for the sum of the residuals. The following device was found particularly useful in reducing this total residual. The total residual in each column was calculated and, the number of points in each column being known, it was possible to make patterns for relaxing the whole column at once. In this way the distribution shown in Figure 24 was obtained. Smaller meshes were used then, as shown in Figure 25, and since the residuals on the boundaries of each mesh were found to change very little, diagonal nets for joining the meshes were unnecessary. Equipotentials and flux lines are shown in Figure 26.

Solution of 3-Dimensional Partial Differential Equations

Although the finite difference equations can be formulated easily for 3-dimensional problems, the number of points involved and the difficulties in recording make the method impracticable in general.

Several special cases of 3-dimensional partial differential equations, however, can be treated in this way. Tranter,[12] for instance, has shown how to reduce the number of independent variables in a given partial differential equation by one by using Fourier transforms, thus reducing 3-dimensional equations to 2-dimensional form. Also, the important class of equations having rotational symmetry can be written in essentially 2-dimensional form.

AXIALLY SYMMETRIC FIELDS

Laplace's equation in this case is

$$\frac{\partial^2 V}{\partial r^2}+\frac{\partial^2 V}{\partial z^2}+\frac{1}{r}\,\frac{\partial V}{\partial r}=0$$

whence we have the equation for the residual

$$\boxed{R_0=V_1+V_2\left(1-\frac{a}{2r}\right)+V_3+V_4\left(1+\frac{a}{2r}\right)-4V_0} \tag{10}$$

where $r\neq 0$. As

$$\lim_{r\to 0}\left(\frac{1}{r}\,\frac{\partial V}{\partial r}\right)=\frac{\partial^2 V}{\partial r^2}$$

One finds for the residual at a point P_0 on the axis

$$\boxed{R_0=V_1+V_3+4V_4-6V_0} \tag{11}$$

where $r=0$.

NUMERICAL EXAMPLES

Cylindrical Air-Gap Choke. Consider the air-gap choke shown in Figure 27. Assuming as boundary conditions a linear drop of magnetic potential drop along AB and CD, and constant potential on the rest of the boundary, relaxation was performed on successively finer nets, as described in the previous section on "Graded Networks." The equipotentials then were plotted and these with the flux lines are shown in Figure 28.

Dynamo-Electric Machines. As shown in a paper by Motz and Worthy[13*] the magnetic field inside a dynamo-electric machine can be calculated by this method. Their result for a 4-pole machine with a slotted armature and distributed windings is shown in Figure 29.

Electron Lenses. Most electron lenses have axial symmetry and hence the governing equation can be written in its 2-dimensional form. Motz and Klanfer[14] have used relaxation to calculate the field distribution in an electrostatic lens, consisting of two semi-infinite cylinders placed end to end and separated by a finite gap. Hesse[15] recently applied the method to the calculation of the field distribution inside a magnetic electron lens, taking account of saturation of the pole pieces.

Solution of Special Partial Differential Equations

TYPES OF EQUATIONS

Difference equations can be formed to replace most differential equations occurring in electrical engineering, and the appropriate relaxation patterns found by similar methods to those already described. If, however, the equation is very complicated, so is the difference equation, and the relaxation pattern may involve a great number of points. It is still possible to work with such patterns as Southwell[16] has shown, but the manipulation then becomes rather involved and tedious. In these cases, however, solution by any other method is practically impossible and so the labor involved in relaxation is justified.

EQUATIONS CONTAINING MIXED DERIVATIVES

The finite difference approximations to mixed derivatives can be found quite easily. With the arrangement of points as shown in Figure 30, a Taylor expansion

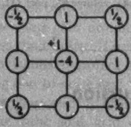


Figure 31 (left). Relaxation pattern for equation

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}+2\,\frac{\partial^2 V}{\partial x\partial y}=0$$

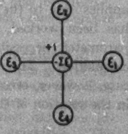


Figure 33 (right). Relaxation pattern for "quasi-plane potential equation"

---

[*] The author wishes to thank Dr. H. Motz and the Institution of Electrical Engineers, London, England, for permission to reproduce Figure 29 from one of his papers.

leads to this approximation

$$4a^2\left(\frac{\partial^2 V}{\partial x \partial y}\right)_0 - V_A - V_B + V_C - V_D$$

The use of this can best be seen from this example

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}+2\frac{\partial^2 V}{\partial x \partial y}=0$$

This can be written approximately

$$\frac{V_1+V_3-2V_0}{a^2}+\frac{V_2+V_4-2V_0}{a^2}+2\frac{V_A-V_B+V_C-V_D}{4a^2}=0$$

whence the residual equation is

$$\boxed{R_0=\Sigma V_1-4V_0+{}^1/_2(V_A-V_B+V_C-V_D)} \qquad (12)$$

The relaxation pattern corresponding to equation 12 is shown in Figure 31.

Quasi-plane Potential Equation

This is the name given to the equation

$$\frac{\partial}{\partial x}\left(\epsilon\frac{\partial V}{\partial x}\right)+\frac{\partial}{\partial y}\left(\epsilon\frac{\partial V}{\partial y}\right)+Z=0$$

which reduces to Poisson's equation, (that is, the plane harmonic equation) when ϵ is constant. It is, for example, the equation for the electric field V in a region of nonuniform dielectric constant ϵ.

With the arrangement of points as in Figure 32, and denoting the value of ϵ between P_0 and P_1 by ϵ_1, and so forth, the finite difference approximation to this equation is found to be

$$\frac{1}{a^2}[\epsilon_1(V_1-V_0)-\epsilon_3(V_0-V_3)+\epsilon_2(V_2-V_0)-\epsilon_4(V_0-V_4)]+Z=0$$

whence the residual formula 13 is obtained

$$R_0=\sum_{n=1}^{4}\epsilon_n V_n - U_0\sum_{m=1}^{4}\epsilon_m+a^2Z \qquad (13)$$

The corresponding relaxation pattern is shown in Figure 33.

Resonator and Oscillation Problems

In most oscillation problems (for example, the field distribution inside a resonator) the governing equation is of the form

$$\nabla^2 V+k^2V=0 \qquad (14)$$

The boundary conditions in waveguide and resonator problems are such that this equation has nontrivial solutions only for certain discrete values of k^2. Solving the equations therefore involves a determination of the allowed parameters k^2 (the "characteristic values") as well as of the corresponding function V. They can be solved by a combination of relaxation and iteration, making use of Rayleigh's principle which gives for the fundamental mode

$$k^2=-\frac{\int\int V\cdot\nabla^2 V dxdy}{\int\int V^2\cdot dxdy} \qquad (15)$$

The procedure is to assume values for V and with these values calculate k^2 from equation 15. This value is substituted in equation 14 which then can be relaxed in the usual way. From this a corrected value of V is found which is used in equation 15 to obtain a corrected value of k^2, and so forth. The process is continued until corrections are negligible, 3 or 4 cycles usually being sufficient.

Allen, Fox, Motz and Southwell[17] applied this treatment to a 2-dimensional section through a klystron tube. Motz[18] extended the method to finding the field distribution and resonance frequency of cavity resonators, improving the accuracy by taking account of the analytic behaviour of fields in the region of sharp corners.

Conclusion

Southwell and his co-workers, who developed the method of relaxation, were interested mainly in applying it to problems in mechanical engineering. However, as this account has shown, the principles and techniques established by them are easily applicable to many problems in electrical engineering. The most important and most frequent use of the method is in the solution of the electromagnetic field problems described by Laplace's, Poisson's, and Maxwell's equations. Its great advantages as compared with analytical methods are:

1. It is simple to apply.
2. The process is essentially the same for all types of equations.
3. It can be checked easily during the progress of the work.
4. It can be applied to problems for which no analytical solution can be obtained.
5. It can mostly be made as accurate as a given problem requires it to be.

The chief disadvantage is that the solution obtained is a numerical one and applies only to the one case for which it has been worked. A variation of any parameter necessitates a complete recalculation. The solution for the first case, however, often gives some guidance in the solution of the next, thus reducing the effort required for the whole range of solutions.

References

1. Relaxation Methods in Engineering Science (book), **R. V. Southwell.** Oxford University Press, London, England, 1940.

2. Relaxation Methods in Theoretical Physics (book), **R. V. Southwell.** Oxford University Press, London, England, 1946.

3. Relaxation Methods Applied to Engineering Problems, Parts 1–11, **R. V. Southwell.** *Proceedings*, Royal Society of London, London, England, 1935, page 56.

4. A Short Account of Relaxation Methods, **L. Fox.** *Quarterly Journal of Mechanics and Applied Mathematics*, Oxford University Press (London, England), volume 1, part 3, 1948, page 253.

5. **H. W. Emmons.** *Quarterly of Applied Mathematics*, Brown University (Providence, R. I.), volume 2, 1944, page 173.

6. **G. F. Freeman.** *Bulletin of Electrical Engineering Education* (London, England), number 2, January 1949, page 7.

7. Approximate Numerical Solution of Electrostatic Fields and Analogous Harmonic Fields, **P. Gazzana Priaroggia.** *L'Energia Elettrica* (Milan, Italy), volume 25, number 10–11 1948, page 533.

8. **M. Salvadori.** *Proceedings*, Seventh International Congress of Applied Mechanics (London, England), September 1948.

9. Some Improvements in the use of Relaxation Methods for the Solution of Ordinary and Partial Differential Equations, **L. Fox.** *Proceedings*, Royal Society of London, (London, England), volume 190, 1947, page 31.

10. Solution by Relaxation Methods of Plane-Potential Problems with Mixed Boundary Conditions, **L. Fox.** *Quarterly of Applied Mathematics*, Brown University (Providence, R. I.), volume 2, 1944, page 3.

11. Plane-Potential Problems Involving Specified Normal Gradients, **R. V. Southwell, G. Vaisey.** *Proceedings*, Royal Society of London (London, England), volume 182, 1943, page 129–51.

12. The Combined use of Relaxation Methods and Fourier Transformers in the Solution of Some Three-Dimensional Boundary Value Problems, **C. J. Tranter.** *Quarterly Journal of Mechanics and Applied Mathematics*, Oxford University Press (London, England), volume 1, part 3, 1948, page 281.

13. Calculation of the Magnetic Field in Dynamo-Electric Machines by Southwell's Relaxation Method, **H. Motz, W. D. Worthy.** *Journal*, Institution of Electrical Engineers (London, England), volume 92, part II, 1945, page 522.

14. Complete Computation of Electron Optical Causes, **H. Motz, L. Klanfer.** *Proceedings of the Physical Society* (London, England), volume 58, 1946, page 30.

15. **M. B. Hesse.** *Proceedings of the Physical Society* (London, England), volume 63B, 1950, page 386.

16. High Speed Flow of Compressible Fluid Through a Two-Dimensional Nozzle, **J. R. Green, R. V. Southwell.** *Philosophical Transactions of the Royal Society of London* (London, England), volume 239, 1944, pages 367–86.

17. **D. N. de G. Allen, L. Fox, H. Motz, R. V. Southwell.** *Philosophical Transactions of the Royal Society of London* (London, England), volume 239, 1945, pages 488–500.

18. **H. Motz.** *Journal*, Institution of Electrical Engineers (London, England), volume 93, part III, 1946, page 21.

No Discussion

Notes on the Design of Eccles-Jordan Flip-Flops

MORRIS RUBINOFF
ASSOCIATE MEMBER AIEE

Synopsis: The design of a prototype Eccles-Jordan flip-flop such as used in large-scale digital computers is discussed, with particular attention to the role played by component variability. A graphical design technique is described; the technique should prove useful in other switching-circuit design problems where a large number of identical direct-coupled circuits are employed.

A preliminary analysis is carried through for the rise time of the on-going control grid. The results may be used to indicate the influence of transpose capacitance on speed of switching of flip-flops.

IT IS common knowledge that the bistable vacuum-tube device referred to as the Eccles-Jordan flip-flop was first described by Eccles and Jordan in 1919.[1] It is perhaps not so well known that the device described by them is not identical with the current flip-flop in that they used batteries in the place of transpose resistors and thus dispensed with resistor divider networks.

The flip-flop now in common use is shown in Figure 1. It is seen to consist of the following: two divider networks each containing a plate resistor R_1, a transpose resistor R_2 usually shunted by a transpose capacitor C_2, and a grid resistor R_3; two tubes, either triodes or pentodes; and two supply voltages, one positive and the other negative with respect to the grounded cathodes. Variations of this configuration are in use; note in particular that it is not necessary that the two halves of the flip-flop be nominally identical. However, this paper will restrict consideration to the illustrated circuit.

Figure 1 illustrates one stable state in which the left tube is completely cut off while the grid voltage of the right tube is slightly above ground. A. T. Starr[2] showed that stability can be attained even though the "off" tube is not cut off and the "conducting" tube has negative grid. In fact, it is easy to show that stability requires only that, in the quiescent state, the loop gain around the two tubes be less than unity. On the other hand, it is advisable to protect against malfunctioning due to noise and, therefore, to reduce the loop gain by ensuring both cutoff for one tube and grid conduction for the other. In any case, it should be noted that a variation in any parameter which increases the grid voltage of the cutoff tube or decreases the grid voltage (or current) of the conducting tube leads the flip-flop closer to instability; this criterion will be used later to determine the influence of parameter variations on flip-flop stability.

Scope of Paper

The design of a single Eccles-Jordan flip-flop for laboratory use is relatively simple and straightforward. On the contrary, the design of a prototype flip-flop for use, for example, in a large-scale digital computer requires more care because of the large number of flip-flops involved. First, the flip-flop must fit into the framework set by other (computer) circuits, which implies specified output voltage levels and switching times. Secondly, the flip-flop must exhibit the desired characteristics in spite of component variations, both initial and during use in the machine. In this respect, attention must be focussed on resistor drift, tube aging, and fluctuations in supply voltages.

The general problem of flip-flop design with all its ramifications is much too complex to permit of its discussion here. Hence, the following assumptions are made in order to restrict and to define the scope of the paper:

1. All resistors will be so selected that after long service in the computer they will not drift outside a given fractional tolerance ρ from nominal resistance.

2. All supply voltages can be held at all times to within a given fractional tolerance σ.

3. Knowledge of tube characteristics for the complete life of every tube (or tube rejection criteria) are such that: 1. plate current i_p for zero grid voltage e_g and specified plate voltage e_p will always exceed a known value, and 2. grid bias for given e_p need never exceed a known value to ensure cutoff.

4. Satisfactory operation is ensured if: 1. tube cutoff is ensured, 2. the grid of the conducting tube reaches at least zero bias, 3. zero grid bias implies zero grid current. (The grid current condition has been included only to simplify exposition. The graphical design procedure described in this paper is readily modified to take account of grid current at zero grid bias.)

5. Only the grounded-cathode flip-flop is to be considered, and no inductors are to be used.

Notation

The notation used in this paper is illustrated in Figure 1. E_1 and E_2 refer to the positive and negative supply voltages, while E and G refer to plate voltage and grid voltage respectively. Algebraic signs have been appended so that all symbols are positive quantities. The subscripts C and NC refer to the C and NC divider networks connected to the conducting and nonconducting anodes respectively. Finally, the transpose capacitance C_2 is assumed to be a variable parameter available to the designer, but the stray capacitances C_1 and C_3 in shunt with R_1 and R_3 respectively are fixed by tube type and physical layout. In this regard it is worth mentioning that C_3 does not include Miller capacitance; this is certainly justifiable for the off tube, which is the only case treated here, but can be largely justified in any case.

Derivation of the "Worst" Configuration

It is proposed now to determine the "worst" possible configuration of the parameters, that is, that configuration most nearly approaching instability as defined above. A graphical approach is to be employed, partly because it has proved to be simpler than the analytic method, but mainly because the graphical method is most illuminating and immediately applicable to design of more-sophisticated flip-flops, of univibrators, and of all circuits using resistance divider networks.

Consider first the NC divider illus-

Paper 52-215, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23-27, 1952. Manuscript submitted December 7, 1951; made available for printing May 12, 1952.

MORRIS RUBINOFF is with the University of Pennsylvania, Philadelphia, Pa.

The author would like to point out that the work reported here was begun at the Institute for Advanced Study Computer Laboratory and to acknowledge incorporation of some of the basic concepts prevalent there. Credit belongs to Mr. P. Levonian for checking some of the lemmas experimentally, to Mr. S. Ruhman for verifying the algebraic steps leading up to the lemmas, and to the clerical staff of the Moore School for their greatly appreciated assistance.

The latter and greater part of the work described here has been done in connection with Signal Corps Contract Number *DA-36-039-sc-14* for Electronic Computer Research and Development, placed by Signal Corps Engineering Laboratories, Department of the Army, Fort Monmouth, N. J.

trated in Figure 2. Note in the upper graph the "minimum" plate characteristic for $e_g=0$, item 3 in the section "Scope of Paper." The divider current i_{NC} is drawn as if it were negative plate current, the reason to be given later. The solid lines labelled R_1, R_2, and R_3 are seen to be "load lines" for the respective resistors, that is, their slopes are the reciprocals of the resistances represented. The broken lines show the effect of more negative-E_2 supply (larger E_2). Clearly, the resultant change in G tends toward flip-flop instability since the originally zero grid goes negative.

Similarly, the lower graph of Figure 2 shows that an increase in R_2 in the NC divider tends toward instability. By graphically studying the effect of varying each parameter, one at a time, its influence on stability is rapidly ascertained.

The C divider can be studied graphically in similar fashion, as shown in Figure 3. Here again the solid lines are load lines. Note in particular that the load line for R_1 fits properly onto this graph only because the divider current i_c has been treated as negative i_p. The graph shows that decreasing R_1 tends to instability because it tends to raise the grid voltage of the nonconducting tube. Actually, it is possible to derive more information from this graph by noting that the change ΔG is due to increased i_c which is in turn due to increase in E_C. It follows that decreasing R_1 will tend to instability whenever the plate characteristic has a positive finite slope at the point (E_C, i_p), a condition generally realized.

The effect of variation of each parameter of the C divider can be studied graphically to yield information regarding tendency toward instability as well as associated conditions on the minimum plate characteristic. Although this is not to be carried through in this paper, the results can be summarized as follows:

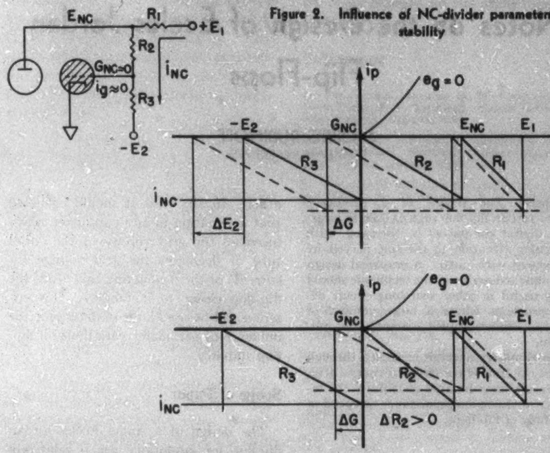


Figure 2. Influence of NC-divider parameters on stability

Lemma I:

(a). For the NC divider, the most nearly unstable configuration is obtained with E_1(min), R_1(max), R_2(max), R_3(min), and E_2(max):

(b). For the C divider, the most nearly unstable configuration is obtained with E_1(max), R_1(min), R_2(min), R_3(max), and E_2(min).

Study of the configurations of lemma I will show that they do not result in extreme values of E; nevertheless, the notation E(max), E(min) will be used for this case.

One might object that I(a) and (b) are inconsistent since E_1 and E_2 are the same for both dividers and cannot be both maximum and minimum simultaneously. However, the purpose is to design a flip-flop stable over long periods of time. Hence, it is necessary to apply both parts of the lemma simultaneously in the initial design. Each part was independently derived for the same reason.

Some Design Lemmas

Lemma I can be applied to derive analytic expressions for the network resistances. Thus from I(a) it follows that

$$i_{NC}=\frac{E_1(\min)}{R_1(\max)+R_2(\max)}=\frac{E_2(\max)}{R_3(\min)} \qquad (1)$$

while from I(b) is obtained

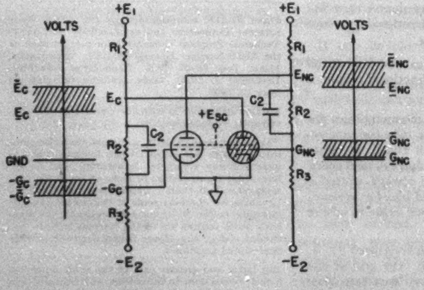


Figure 1. Grounded-cathode Eccles-Jordan flip-flop. (Superscript and subscript bars denote maximum and minimum respectively

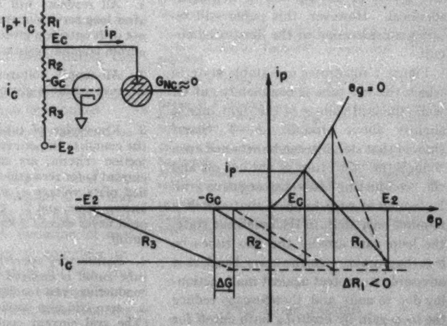


Figure 3. Influence of C-divider parameters on stability

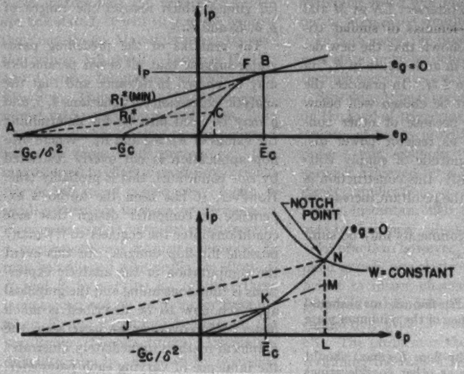


Figure 4 (left). Determination of optimum E_C(max). (Superscript and subscript bars denote maximum and minimum respectively)

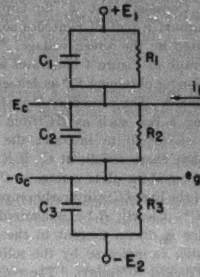


Figure 6 (right). Initial conditions for rising grid voltage

$$i_c = \frac{E_c(\max)+G_c(\min)}{R_3(\min)}$$
$$= \frac{E_2(\min)-G_c(\min)}{R_3(\max)} \qquad (2)$$

$$i_p + i_c = \frac{E_1(\max)-E_c(\max)}{R_1(\min)} \qquad (3)$$

These equations can be transferred by straightforward algebraic manipulation to

$$i_pR_1(\min) = E_1(\max)\delta^2 + G \qquad (4)$$

$$i_pR_2(\min) = \frac{E+G}{1-g} \times$$
$$\left[\frac{1}{\beta^2 - \dfrac{E+G}{(E_1(\max)-E)(1-g)}} - (1-g)\right] \qquad (5)$$

$$i_pR_3(\max) = E_2(\min) \times$$
$$\left[\frac{1}{\beta^2 - \dfrac{E+G}{(E_1(\max)-E)(1-g)}} - (1-g)\right] \qquad (6)$$

$$\beta^2 = \frac{E_1(\max)}{E_1(\max)-E}\theta^2 \quad \delta^2 = 1-\theta^2(1-g)$$

$$\theta = \frac{(1-\rho)(1-\sigma)}{(1+\rho)(1+\sigma)} \qquad g = \frac{G}{E_2(\min)}$$

$$E \equiv E_c(\max) \qquad G \equiv G_c(\min)$$

The following lemmas may be seen to follow almost immediately from equations 4, 5, and 6:

Lemma II:

R_1, R_2, and R_3 are all inversely proportional to i_p; that is, all other things equal, the resistances are minimized by choosing a tube with maximum perveance. (This lemma is modified slightly when $i_g \neq 0$ for $e_g = 0$.)

Lemma III:

R_1 is increased, R_2 and R_3 are decreased when a larger value of E_1 is chosen.

Lemma IV;

If E_2 be chosen appreciably larger than G so that g is small, then R_1 and R_2 are almost independent of E_2. R_3 is approximately proportional to E_2.

Lemma IV(a):

A reasonable choice for E_2 is $E_2 \cong 10G_c(\min]$.

Lemma V:

All three resistors increased as either resistance tolerance ρ increases or supply tolerance σ increases or both, for then θ and β decrease and δ increases.

Since neither R_2 nor R_3 can be negative, it is necessary for the denominator of the first term inside the bracket of equations 5 and 6 to be positive. That is

$$\beta^2 - \frac{E+G}{(E_1(\max)-E)(1-g)} > 0 \qquad (7)$$

and therefore

Lemma VI:

The minimum permissible value of E_1 is limited by the inequality

$$E_1(\max) > \frac{E+G}{1-g}\cdot\frac{1}{\theta^2}$$

Finally it is proposed to derive a criterion for choice of E_C(max). The criterion will be based on the assumption, believed to be approximately true, that flip-flop speed is inversely proportional to R_1, that is, that R_1 should be minimized. By lemma III, this occurs when E_1 is chosen as small as possible. Applying lemma VI to equation 4, one obtains

$$i_pR_1(\min) > \frac{(E+G)\delta^2}{(1-g)\theta^2} + G = \frac{\delta^2}{1-\delta^2}\left(E + \frac{G}{\delta^2}\right) \qquad (8)$$

It is helpful to introduce the resistance R_1^* given by

$$i_pR_1^* = E + \frac{G}{\delta^2} \qquad (9)$$

Thus, R_1^* is a convenient measure of R_1.

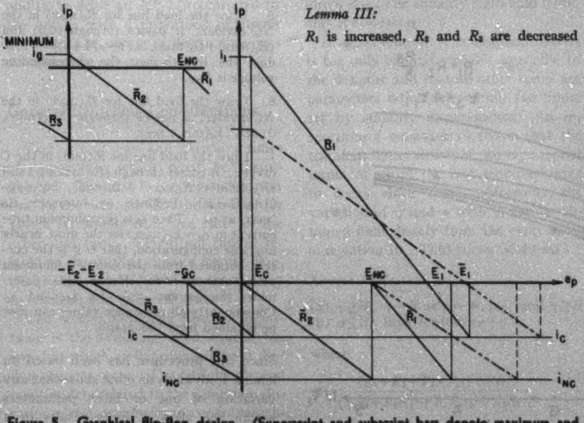


Figure 5. Graphical flip-flop design. (Superscript and subscript bars denote maximum and minimum respectively)

Consider first a flip-flop designed around a pair of pentodes so that G is fixed by the screen voltage. The upper graph of Figure 4 illustrates a construction for obtaining $R_1{}^*$ as defined by equation 9. The procedure consists of drawing a line such as AC from the point $(-G/(\delta^2),0)$ to intersect the minimum plate characteristic at C; if E be chosen to equal the plate voltage at C, the slope of the line AC is inversely proportional to $R_1{}^*$. Clearly, $R_1{}^*$ is minimized when the line is drawn tangent to the characteristic, as indicated by the solid line AB (labelled $R_1{}^*(\text{min})$). The point of tangency determines the optimum value of E. Since $0<\delta<1$, $R_1{}^*$ can never be decreased below the value corresponding to the dashed line passing through $(-G_C(\text{min}), 0)$ and tangent at F. Clearly, E should be chosen somewhere on the "knee" of the curve.

In the case of a triode flip-flop, the same reasoning leads to the result that E be chosen as large as possible. The upper limit is set by the rated anode power dissipation W. Hence, E should be chosen at the "notch point" N at which the minimum plate characteristic intersects the "rated power" curve W=constant as shown in the lower graph of Figure 4. The broken line IN indicates the corresponding $R_1{}^*(\text{min})$.

It should be recalled, however, that in the case of a triode, G is roughly proportional to E_1 which is in turn roughly proportional to E. Hence $R_1{}^*$ does not increase with decreasing E as rapidly as might otherwise be anticipated. The solid line JK shows the effect of reducing all voltages by a factor of two. By drawing a line from the origin through the new design point K to intersect LN at M and by applying the lemmas of similar triangles, it can be shown that the new design point results in an increase in $R_1{}^*$ in the ratio of LN to LM. In practice, the design point must be chosen well below the notch point because of other considerations, such as resistor power dissipation and magnitude of supply voltages. Consequently, this construction is a useful guide to the resultant increase in R_1.

The preceding comments may be summarized as follows:

Lemma VII:

(a) For a pentode flip-flop, $E_C(\text{max})$ should be chosen on the knee of the minimum plate characteristic.

(b) For a triode flip-flop, $E_C(\text{max})$ should be chosen as large as other considerations permit.

Graphical Design Procedure

The analytic expressions for the network resistances given by equations 4, 5, and 6 of the section entitled "Some Design Lemmas" contain the following seven parameters, E_1, E_2, E, G, i_p, ρ, σ. The magnitudes of these parameters may be assigned arbitrarily by the designer, subject only to the condition on E_1 stated in lemma VI. The assignment of values for ρ and σ is dependent upon knowledge of the characteristics of the resistors and of the power supplies to be used, while the selection of G and i_p depends upon a knowledge of the characteristics of existing vacuum tube types; clearly, engineering judgment must enter into the selection of values for these. At the same time, lemmas IV, V and VII serve as useful guides which restrict the ranges of ρ, σ, E_2 and E.

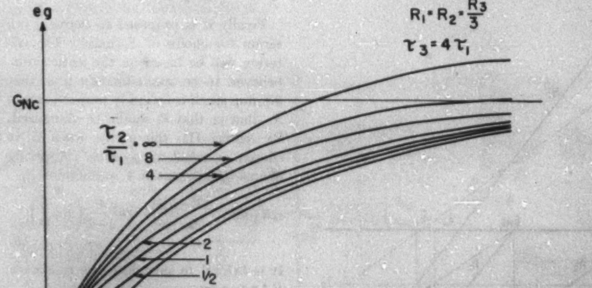


Figure 7. The influence of τ_2 on switching time

The remarks of the preceding paragraph suggest that all seven parameters may be chosen in advance and that the analytic expressions of equations 4, 5, and 6 may be used directly for determining the values of R_1, R_2, and R_3. Where flip-flop application is not overly restricted by side conditions, this is probably true. However, it has been the author's experience in computer design that side conditions force the engineer to try many possible flip-flop designs. In this event the computation of the analytic expressions is time consuming and the graphical approach now to be described is much faster. Moreover, as we have seen, the graphical method immediately illustrates the influence of varying each parameter.

The design procedure is based directly upon lemma I and the assumptions specified in the section entitled "Scope of the Paper." The procedure is illustrated in Figure 5.

1. Set up i_p, e_p axes; the voltage scale is known but the current scale is unknown initially.

2. Draw a line parallel to and below the e_p axis to represent i_c, the C-divider current. (It is convenient to make $i_c=10$ divisions.)

3. Locate $-E_2(\text{min})$, $-G_C(\text{min})$, $E_C(\text{max})$. Optimum choices for E_2 and E are based on lemmas IV(a) and VII; $-G_C(\text{min})$ is the guaranteed grid cutoff bias. Draw load lines for $R_2(\text{min})$ and $R_3(\text{max})$.

4. Draw the load line for $R_3(\text{min})$ in the NC divider. This line passes through $(-E_2(\text{max}), 0)$ and $(-G_C(\text{min})-[E_2(\text{max}) -E_2(\text{min})], i_c[1+\rho/1-\rho])$. Produce this line to cut the current axis; the intersection is at i_{NC}, the current through all three NC-divider resistors.

5. Draw the load line for $R_2(\text{max})$ in the NC divider; it passes through (0,0) and $[E_c(\text{max})+G_c(\text{min}),\ i_c(1-\rho/1+\rho)]$. Produce the line to i_{NC}; the corresponding voltage is E_{NC}.

6. Draw the load line for $R_1(\text{max})$ in the NC divider; it passes through $(E_{NC}(\text{min}), 0)$ and $(E_1(\text{min}), i_{NC})$.

7. Draw the load line for $R_1(\text{min})$ in the C divider; it passes through $(E_1(\text{max}), i_c)$ and $[E_{NC}(\text{min})+ E_1(\text{max}) - E_1(\text{min}),\quad i_c - i_{NC} (1+\rho/1-\rho)]$. Produce to intersect E_C (max) at i_p. Then i_p is the minimum tube current at $e_p=E_C(\text{max})$ for the most nearly unstable configuration, that is it is the current obtained from the assumed minimum plate characteristic. Hence this determines the current scale for i_{NC} and i_c. Consequently all resistance values can now be obtained from this graph.

Since this procedure has been based on lemma I, an analysis must show that any variation of one or more parameters within the allowed limits results in a stable flip-flop configuration.

No matter how the resistance values are calculated, it will in general happen that non-Radio Manufacturers Association values will result. Lemmas II, III, and IV may be applied to rectify this situation. Thus, since i_p has been determined using engineering judgment, it is quite reasonable to alter the preassumed value by some small percentage; this may be used to effect a small proportional rise or fall in all resistances. Again, a small percentage increase or decrease in E_1 may be used to vary the resistances in a nonproportional manner. Finally, R_3 may be adjusted with negligible effect on R_1 and R_2 by varying E_2. (Where these lemmas can not be used, as for example when E_1 and E_2 cannot be altered, the non-Radio Manufacturers Association difficulty may be alleviated by designing with larger tolerances, ρ and σ, than engineering judgment demands.)

The power of the graphical method is illustrated by two remarks. First, the broken lines on the right-hand side of Figure 5 instantly depict the influence of increased E_1 on the divider resistors. Since i_p is given and does not change, larger E_1 results in larger i_{NC} and i_c and hence in smaller R_2 and R_3; at the same time, R_1 increases as is evident from inspection of the intersection of the two load lines for R_1(min) with the voltage axis. Second, the graphical approach is easily extended to take account of grid current at zero grid volts; the method is indicated by the insert top left of Figure 5. The author has used the graphical method for a case when the analytic expressions are particularly complex; this was the design of a flip-flop with a guaranteed minimum grid current and also the ability to drive an *NE-17*, connected to the anode, with a prescribed minimum current. The saving in design time was deemed to be substantial.

General Switching Considerations

There are at least two switching times of importance which must be distinguished. The first is the repetition rate that is, the rate at which the flip-flop can switch repeatedly from one state to the other and back. Repetition rate is related to the settling time of the flip-flop, that is, the time required for the flip-flop to reach its new quiescent state.

The second switching time of importance is the switching speed, that is, the speed at which an initially quiescent flip-flop changes state. The switching speed is intimately related to the minimum duration of trigger pulse for successful switching. It will be shown that in general switching speed and repetition rate cannot be maximized simultaneously.

In both cases, the time of arrival at the new state must be defined *a priori*, say, through specification of output voltage.

Switching time is necessarily dependent on the kind of triggering used and on the amount of energy provided by the trigger pulse. For example, the ultimate in speed is achieved by a trigger which, by direct application to all appropriate points, instantaneously switches all voltages to their new values.

In practice, however, the trigger is applied to only one electrode for a time sufficient to insure change of state. It is proposed to consider the case when the conducting grid is instantaneously cut off. The problem is then reduced to determining the rate of rise of grid voltage of the initially nonconducting tube in response to a step of current, i_1, injected into the C divider between R_1 and R_3, as indicated in Figure 6. It is easily shown that

$$e_g(t)=G_{NC}-(G_{NC}+G_C)\times[(1+b)\epsilon^{-\alpha t}-b\epsilon^{-\gamma t}] \quad (10)$$

where

$$b=b(\tau_2)=\frac{\alpha(1-\gamma\tau_2)}{(\gamma-\alpha)}$$

$$\gamma,\alpha=\frac{1}{2R_i\tau_j\tau_k}[(R_i+R_j)\tau_k\pm S]$$

$$S^2=[(R_i+R_j)\tau_k]^2-4(R_i)(R_i\tau_j\tau_k)$$
$$\tau_k=R_kC_k$$
$$R_i=R_1+R_2+R_3$$
$$R_i\tau_j\tau_k=R_1\tau_2\tau_3+R_2\tau_3\tau_1+R_3\tau_1\tau_2$$
$$(R_i+R_j)\tau_k=(R_1+R_2)\tau_3+(R_2+R_3)\tau_1+(R_3+R_1)\tau_2$$

Also it can be shown that $S^2\geqslant 0$ and therefore $\gamma\geqslant\alpha$. The equality signs hold if and only if $\tau_1=\tau_2=\tau_3$.

In what follows, it is assumed that τ_2 is the only explicit variable available to the designer and that all other terms are parameters adjustable within the limits set by stability requirements. An inspection of equation 10 shows that the transient term consists of two exponential terms, of which the first requires longer to decay since $\alpha<\gamma$. Moreover, the variation of γ and α with τ_2 can be obtained immediately from the expressions of equation 10. Thus it can be shown

Lemma VIII:

(a) $\gamma\geq H$; $\gamma\rightarrow H$ as $\tau_2\rightarrow\infty$, but limit $\gamma\geq H$
(b) $\alpha\leq H$; limit $\alpha=0$ as $\tau_2\rightarrow\infty$

where

$$H=\frac{(R_1+R_2+R_3)(R_1\tau_3+R_3\tau_1)}{(R_1\tau_3+R_3\tau_1)^2+R_1R_2\tau_3^2+R_2R_3\tau_1}$$

When $\tau_1=\tau_3$, then $H=1/\tau_1$ and the equality signs in lemma VIII are obeyed. Moreover, note that the largest value of α is less than the smallest value of γ, regardless of the value of τ_2, for any specific design.

The coefficient b has been defined in equation 10; in particular, it can be seen that

$$b(0)=\frac{\alpha}{\gamma-\alpha}>0 \quad (11)$$

It follows that both coefficients are positive for $\tau_2=0$. It can be shown that:

Lemma IX:

As τ_2 increases, both coefficients decrease through zero and become negative. In particular, $b=0$ when τ_2 equals the smaller of τ_1 and τ_3; $(1+b)$ equals zero when τ_2 equals the larger of τ_1 and τ_3.

Since α can never exceed γ, minimum settling time can be achieved only by eliminating the term exponential $(-\alpha t)$, that is, by making $(1+b)=0$. Thus

Lemma X:

To achieve minimum settling time, τ_2 should be chosen equal to the larger of τ_1 and τ_3. Settling time is then dependent only upon γ.

Two observations are worthy of note. First, this value of γ does not differ too markedly from $1/\tau_1$. It follows that the magnitude of τ_1 constitutes a good first approximation to flip-flop speed. This circumstance was utilized in the section entitled "Graphical Design Procedure" in choosing optimum E_c(max) on the basis of minimum R_1. Second, when τ_2 equals the larger of τ_1 and τ_3, the equation for γ simplifies to

$$\gamma=\frac{R_1+R_2+R_3}{(R_1+R_2)\tau_3+R_3\tau_1};\quad \tau_1=\tau_2>\tau_3$$
$$\gamma=\frac{R_1+R_2+R_3}{R_1\tau_3+(R_2+R_3)\tau_1};\quad \tau_1<\tau_2=\tau_3 \quad (12)$$

As pointed out, equation 12 was chosen to satisfy minimum settling time. It is common experience, however, that repetition rate can be improved by introducing a small amount of overshoot at the expense of settling time. It is not difficult to show the following:

Lemma XI:

Overshoot occurs if and only if τ_2 is greater than both τ_1 and τ_3.

It has been found experimentally for one flip-flop design that maximum repetition rate was obtained when τ_2 exceeded the larger of τ_1 and τ_3 by about 10 per cent.

To obtain maximum switching speed requires that the grid voltage e_g rise at its maximum rate. By differentiating e_g with respect to time t in equation 10, it can be shown that:

Lemma XII:

For maximum switching speed, τ_2 should be so chosen that

$$\tau_2 >> \frac{R_3\tau_3\tau_1}{R_1\tau_3+R_3\tau_1} \text{ or } C_2 >> \frac{C_1C_3}{C_1+C_3}$$

In words, C_2 should be chosen much larger than the capacitance of the series connection of C_1 and C_3.

The influence of τ_2 on switching speed is illustrated by the curves of Figure 7, which show e_g versus time for various values of τ_2. Note that these curves are pictorial verification of lemmas X, XI, and XII.

The use of large τ_2 to obtain maximum switching speed implies the existence of appreciable overshoot. In the case of the rising grid, this is of no consequence since the grid is clamped at cathode voltage. On the other hand, it is the same network which is involved in fall time considerations when the flip-flop returns to its original state. Consequently, there will be a similar overshoot in the case of the falling grid, probably of increased magnitude because of the initial grid clamping. The overshoot reduces the ultimate repetition rate since it causes the grid, when retriggered soon after having been triggered, to start out from a voltage below $-G_C$ and hence to take longer to rise. Thus it appears qualitatively that the two switching times, repetition rate and switching speed, cannot be optimized simultaneously.

Conclusion

This paper has been concerned with the design of Eccles-Jordan flip-flops. After restricting the scope of the discussion to a particular design and to specific assumptions, analytic design expressions were obtained and a number of lemmas developed with regard to optimizing parameters while guaranteeing stability within stated parameter tolerances. Finally, a general graphical design procedure was developed.

The use of a transpose capacitor to improve switching times was discussed and a number of lemmas developed to serve as a guide not only in the choice of transpose condenser but also in choice of other parameters (particularly R_1 and R_2) within the stability limits previously obtained.

The paper makes no claim to completeness but is meant only to serve as a collection of rules to guide the design of flip-flops.

References

1. A Trigger Relay Utilizing 3-Electrode Thermionic Vacuum Tubes, W. H. Eccles, F. W. Jordan. *Radio Review* (Dublin, Ireland), volume 1, 1919, pages 143–46.

2. A Trigger Peak Voltmeter Using "Hard" Valves, A. T. Starr. *Wireless Engineer* (London, England), volume 12, number 146, 1935, pages 601–06.

No Discussion

Ultrathin Tapes of Magnetic Alloys with Rectangular Hysteresis Loops

MARTIN F. LITTMANN
NONMEMBER AIEE

Synopsis: The principal soft magnetic alloys have been rolled to thicknesses between 1/8 and 1 mil. Magnetic properties of toroidal cores wound from ultrathin tapes were studied using d-c magnetization. Even though the coercive force tends to increase with reduced thickness, the coercivity of ultrathin metal tapes is usually lower than that of ferrites or powdered metals. Relatively rectangular d-c hysteresis loops were obtained in three of four alloys investigated.

THE continuing need for magnetic core materials which can be used effectively at high frequencies has stimulated the search for materials in which eddy-current effects produced by rapid changes in magnetic flux are at a minimum. Generally speaking there are two ways in which eddy-current effects can be reduced: increasing the electrical resistivity as in the ferrites, or restricting the paths in which eddy-currents can flow, as in powdered metals or thin laminations of metal strip.

The ferrites, which have resistivities about 10^6 to 10^{12} higher than conventional magnetic alloys, have been used extensively in television sweep transformers where they may be operated at high frequency with low power loss.[1] However, such characteristics as relatively low magnetic saturation, low Curie point and certain dielectric properties, particularly for larger masses, restrict the scope of application of the ferrites. On the other hand, the fine subdivision of particles in powdered metal

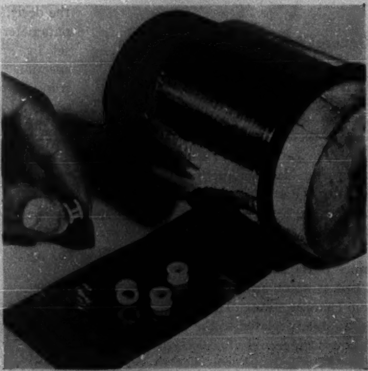


Figure 1. Coil of cold rolled 48-per-cent nickel-iron alloy 1/4-mil thick and ceramic bobbins

---

Paper 52-208, recommended by the AIEE Magnetic Amplifiers and Basic Sciences Committees and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted February 7, 1952; made available for printing May 9, 1952.

Martin F. Littmann is with Armco Steel Corporation, Middletown, Ohio.

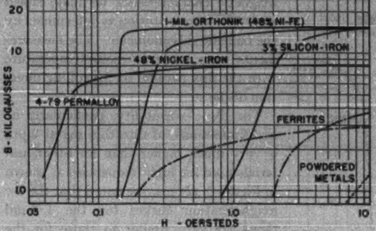


Figure 2 (above). D-c magnetization curves of 1/4-mil ultrathin metal tapes compared with other materials used at high frequencies

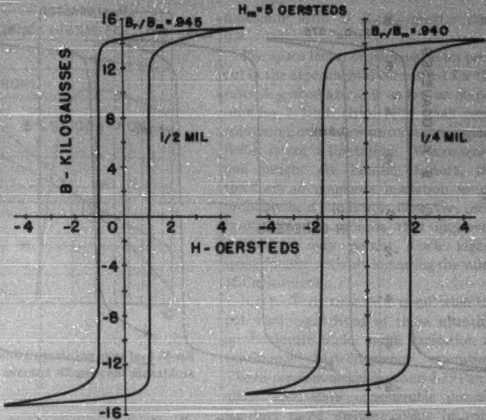


Figure 3 (right). D-c hysteresis loops of grain-oriented 3-per-cent silicon-iron

cores results in very low d-c permeability even for the best alloys. Where high permeability is desirable, very thin metal tapes are useful and these are the basis of the present paper.

The commercial use of magnetic alloys rolled to thin gauges grew very rapidly during World War II. For power transformers, laminated or tape-wound cores of strip 4- to 6-mils thick were employed, and for radar pulse transformers cores of 1- to 3-mil-thick materials were used. This large-scale use of thin materials was, to a large degree, the result of development of economical methods of rolling 1- to 6-mil strip, primarily silicon-iron, in quantities of thousands of pounds. Already during this period there were some applications involving rates of magnetization so high that materials even thinner than 1 mil (0.025 millimeter) were required. Small quantities of *4-79* Permalloy tapes 1/4-mil thick were rolled on a small mill at the Bell Telephone Laboratories for such applications.[2] The production of magnetic alloys in tapes as thin as 1/8 mil has now been undertaken by commercial organizations.

Since 1950, four different alloys have been rolled to gauges less than 1 mil and fabricated into toroidal cores. The descriptive term ultrathin has been applied to these materials. The alloys chosen for study are commonly used soft magnetic materials having high permeability. The group was restricted to those alloys having relatively high resistivity in order to keep eddy-current effects to a minimum. The values of magnetic saturation, Curie temperature, and volume resistivity, as shown in Table I, are characteristic of the alloys themselves and probably do not change with thickness in the range of 1 mil to 1/8 mil.

The alloys were rolled as single strands in a width of 2 inches employing work rolls of very small diameter. After rolling, the strip was slit into ribbons 1/8-inch or 1/4-inch wide for winding into toroidal cores. Because of the fragile nature of small toroidal cores, often consisting of only a few wraps of tape, it was advantageous to wind the cores on ceramic bobbins as shown in the foreground

Table I. Basic Properties of Alloys Investigated

	3-Per-Cent Si, Balance Fe	48-Per-Cent Ni, Balance Fe	79-Per-Cent Ni, 4-Per-Cent Mo, Balance Fe	79-Per-Cent Ni, 5-Per-Cent Mo, Balance Fe
Magnetic Saturation, Gausses	20,000	16,000	8,700	8,000+
Curie Temperature, Degrees Centigrade	700	440	420	400
Resistivity, Microhm-Centimeters	47	45	55	60
Density, Grams per Cubic Centimeter	7.65	8.25	8.72	8.77

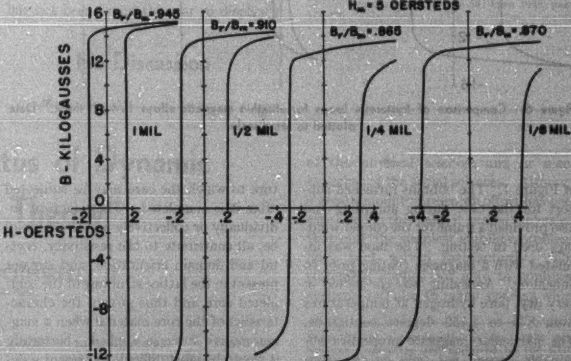


Figure 4. D-c hysteresis loops of grain-oriented 48-per-cent nickel-iron

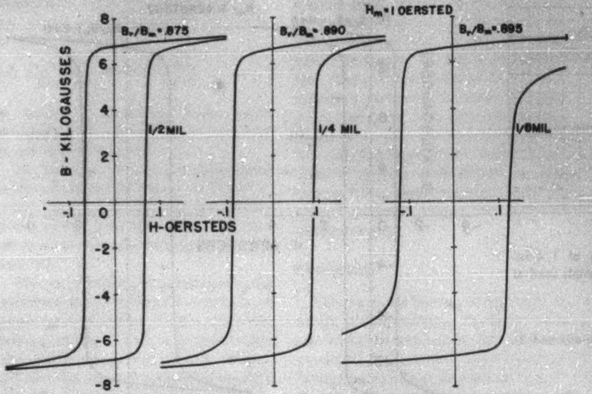


Figure 5. D-c hysteresis loops of *4-79* Permalloy showing pronounced rectangularity

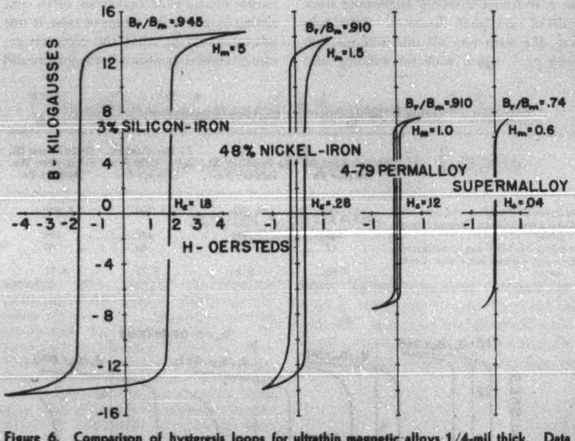


Figure 6. Comparison of hysteresis loops for ultrathin magnetic alloys 1/4-mil thick. Data plotted to same scale

of Figure 1. The bobbins furnished support for the tape during annealing and also provided a frame for the copper windings used in testing. The tape was insulated with a magnesia coating prior to annealing. Annealing was conducted in very dry, pure hydrogen at temperatures from 875 to 1,150 degrees centigrade. The particular magnetic properties obtainable depend not only on the composition of the alloy but on the processing steps of the metal supplier, the forming and annealing procedure used by the fabricator, and the stresses and temperature to which the core may be subjected after it is completed. These factors, individually or collectively as the case may be, all contribute to the resistivity, crystal and domain orientation, and stresses present in the lattice structure of the completed core and thus govern the characteristics of the core material when a magnetomotive force is applied. Naturally it would be most difficult to present a detailed picture of the magnetic properties of ultrathin magnetic alloys at this early stage of development. Our purpose is to present a general picture of the type of magnetic characteristics we may expect under d-c magnetization as determined by the ballistic galvanometer. The measurements were made at room temperature with the flux parallel to the rolling direction of the strip.

The d-c permeability, particularly at high inductions, of ultrathin metal tapes is very much higher than for other materials used for high frequencies as shown in Figure 2. It may be noted that the d-c magnetization curves for the 1/4-mil tapes are very similar to those for the same alloys in the more familiar thicknesses of 1 to 14 mils. For comparison, the magnetization curve is shown for a specimen of 1-mil, grain-oriented 48-per-cent nickel-iron having a very rectangular hysteresis loop. Rectangular hysteresis loops are of considerable interest for applications such as magnetic amplifiers and memory cores and, as will be shown, ultrathin metal tapes may be made to have hysteresis loops which are quite rectangular.

For example, the loops shown in Figure 3 for 1/2-mil and 1/4-mil grain-oriented* 3-per-cent silicon-iron show a high degree of rectangularity. Comparing the width of the two loops it may be noted that the coercive force is almost doubled by the reduction in thickness from 1/2 to 1/4 mil. At first thought this might be considered to be the result of an increased volume effect of surface contamination during final annealing. Of course, these very thin tapes are particularly sensitive to damage from faulty atmosphere during annealing. However, the increase in coercive force with reduced thickness has been observed even under very carefully controlled annealing conditions. The suggestion is made that as thickness is reduced there are changes in domain size and shape which cause an increase in coercive force. A contributing factor may be the type of crystal orientation present, particularly for material such as the 3-per-cent silicon-iron, which has a high coefficient of anisotropy.

As shown in Figure 4, the coercive force for grain-oriented† 48-per-cent nickel-iron also increases with reduced thickness but not as rapidly as for the silicon-iron. Compared to the 3-per-cent silicon-iron the coercive force for 48-per-cent nickel-iron is roughly 10 times lower for thicknesses below 1 mil.

4-79 Permalloy is generally regarded as a material having high permeability particularly at low inductions. Hence, the

---

*Substantially that of a single crystal with a cube edge in the rolling direction and a face diagonal in the transverse direction.

†Substantially that of a single crystal with a cube edge both in the rolling and transverse directions.

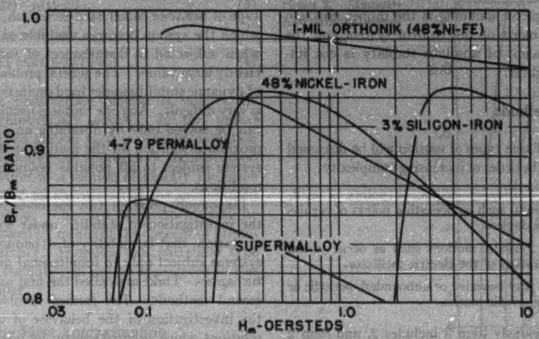


Figure 7. Variation of ratio of residual to maximum induction with magnetizing force for 1/4-mil cores with rectangular hysteresis loops. Data for 1-mil grain-oriented 48-per-cent nickel-iron included for comparison

high degree of rectangularity of the hysteresis loop which it was possible to obtain in ultrathin gauges of *4-79* Permalloy was hardly expected, see Figure 5.

The hysteresis loops for 1/4-mil experimental specimens of the various alloys including Supermalloy are compared with each other on the same scale in Figure 6 which illustrates how the materials compare in a qualitative way. The comparison is intended to be on a qualitative basis since it will be understood that the properties for each alloy can be varied considerably and the particular specimens, data for which are presented here, were chosen because their hysteresis loops were quite rectangular. Again referring to Figure 6, the various shapes of loops illustrate that the terms "square loop" or rectangular hysteresis loop are, at best, only descriptive and that important differences in degree exist depending on the criteria used. Such criteria may be the slopes of the sides of the loops, or the slope of the top of the loop as saturation is approached.

One useful way of measuring rectangularity is to note how the ratio of residual to peak or maximum induction B_r/B_m varies with peak or maximum magnetizing force H_m. The curves for H_m versus B_r/B_m shown in Figure 7 represent the same cores used for Figure 6 and Figure 2. Lower ratios of B_r/B_m may be achieved by variation of processing and annealing treatments. Very low ratios may be obtained by the introduction of a small air gap into the magnetic path. Supermalloy is the only material among those studied here in 1/4-mil thickness which has not been found thus far to display a maximum ratio of B_r/B_m greater than 0.9.

The space factor or packing factor present in the experimental cores used for the tests described above is as low as 50 per cent for 1/4-mil strip. For small cores containing only a few turns of strip, space factor is not a limitation. Where space and weight are limiting factors, the thickness of magnesia insulation would probably be a handicap. However, there is considerable promise that improved techniques may permit much higher space factors without impairing the value of the insulation.

It is too early to state in detail what the potential capabilities of these ultrathin materials are under pulse excitation or continuous high-frequency operation. The 48-per-cent nickel-iron and *4-79* Permalloy tapes show considerable promise for high-speed memory cores in computers. For magnetic amplifiers and radar pulse transformers, the low coercive force and high magnetic saturation of ultrathin alloy tapes are of interest, although it may not be desirable necessarily to have a rectangular d-c hysteresis loop. Ultra-thin *4-79* Permalloy appears to be particularly attractive since it has relatively high volume resistivity and can be made to have either a rectangular or rounded d-c hysteresis loop together with low coercive force.

References

1. Ferrite Materials Permit Improved Designs of Magnetic Devices, Eberhard Both. *Materials and Methods* (New York, N. Y.), volume 34, July, 1951, pages 76–79.

2. Applications of Thin Permalloy Tape in Wide-Band Telephone and Pulse Transformers, A. G. Ganz. *Electrical Engineering (AIEE Transactions)*, volume 65, April 1946, pages 177–83.

No Discussion

The Current Status of Dynamic Stability Theory

FRANK E. BOTHWELL
MEMBER AIEE

ALTHOUGH there exists today no theory of dynamic stability of sufficient generality to satisfy the demands of engineering, there are, nevertheless, several tools which may be employed advantageously by the engineer in many stability investigations. A short description of these tools illustrated by certain practical examples will be presented.

The concepts of absolute and relative stabilities of dynamical systems are matters of much concern in nearly all branches of the physical sciences and in many branches of the social sciences. Although a great amount of research effort has been expended in this field, it can be said safely that only the simplest stability problems have been solved satisfactorily. This statement applies not only to nonlinear systems, but also to many systems which

---

Paper 52-214, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted July 19, 1951; made available for printing May 9, 1952.

Frank E. Bothwell is with the United States Naval Ordnance Test Station, China Lake, Calif.

can be represented in their ranges of operation by entirely linear equations, as, for example, the linear differential equation with periodic coefficients.

The fundamental problem is the definition of stability. It is difficult to find one definition which will suit all purposes. Asymptotic stability, as formulated by Liapounoff,[1] is a satisfactory tool for the investigation of the action of a regulator or a feedback amplifier but is unsatisfactory in the treatment of conservative systems, unless all periodic behavior is termed unstable. Liapounoff's criterion of stability of so-called steady motion rectifies this difficulty for cases (such as linear conservative systems) in which the period of oscillation is independent of amplitude. However, such a criterion applied to the motion of celestial bodies would produce the embarassing conclusion that the motion of each planet is one of unstable equilibrium. Thus it is necessary to introduce a third concept; namely orbital stability.[2]

The three concepts, asymptotic stability, orbital stability, and stability in the sense of Liapounoff (hereafter to be referred to as temporal stability) may be termed absolute stabilities. They do not provide, in themselves, a solution to all problems of definition. A servomechanism, although absolutely and asymptotically stable, may exhibit such a persistent oscillatory response as to render its usefulness no greater than if it had been unstable. Another concept must be introduced, therefore, and one is concerned with degree of damping or relative stability.

Although too little damping is objectionable in many systems, in a few cases zero damping or even a slight negative damping is permissible. Consider the case of the guided missile, the flight path of which is often characterized by oscillations occasionally increasing in amplitude. Inasmuch as the total flight time of the missile may include no more than three or four oscillations, the magnitude of the negative damping is the proper criterion of merit of the control system and the concept of absolute stability is not important.

Methods

Three types of motion will be discussed in the order of increasing complexity:

1. Stationary points, or points of equilibrium, such as operating points of regulatory devices.
2. Periodic motions such as occur in the operation of the electric oscillator.
3. Any bounded or unbounded, periodic or nonperiodic motion.

Obviously item 3 includes 2, and item 2 includes 1. However, items 1 and 2 are more commonly encountered than is item 3, and there is some advantage in a separate treatment.

The particular motion, the stability of which is under investigation, whether it be an equilibrium point or a periodic or nonperiodic wave form, is generally denoted as the undisturbed motion, whereas any other motion of the dynamical system is called a disturbed motion. Broadly speaking, a system is said to be asymptotically stable if all disturbed motions sooner or later degenerate into the undisturbed motion itself. Application of the phrase "all disturbed motions" must be made with reservation however. Consider the case of the periodic motion of a wrist watch. Any light shock will change its period momentarily, but eventually the motion reverts to its former value, and consequently the timepiece is labeled stable. However, if dropped from a second story window, the watch quickly reaches a condition of instability.

The situation is exemplified by the electric oscillator which will not oscillate unless first dropped on the floor, or perhaps by the more practical case of the amplifier which "sings" when overloaded and continues in this state after removal of the excitation. These two examples refer to a type of behavior called hard excitation, in which the system becomes unstable only when subjected to disturbances of sufficiently large value. The general problem of dynamic stability under hard excitation is very difficult. Any one problem usually can be solved only by obtaining an approximate solution for the motion of the system subject to all possible boundary conditions.

Liapounoff presents two methods for the investigation of stability under soft excitation, that is, stability of dynamical systems subject only to infinitesimal disturbances. These are called the first and second methods. The former involves the investigation of the behavior of all solutions which originate in an infinitesimal region surrounding the undisturbed motion. The second method involves the investigation of the existence of a function which, in connection with the given dynamical equations, possesses a certain prescribed behavior. The difficulty in finding such restricted functions has resulted in very little use of the second method, and consequently stability investigations are effected by means of the first method.

Definition

Consider a dynamical system described by the set of differential equations

$$\frac{dx_i}{dt}=f_i(t,x_1,x_2,\ldots,x_n),\quad i=1,2,\ldots,n, \tag{1}$$

in which the functions f_i are, in general, nonlinear. Let $x_{i0}(t)$ denote a particular undisturbed motion, and let $x_i(t)$ denote any disturbed motion. The undisturbed motion is temporally stable if, and only if, for a given positive number ϵ, no matter how small, it is possible to find another number δ such that for all disturbed motions satisfying the initial inequality

$$|x_{i0}(0)-x_i(0)|<\delta \tag{2}$$

the subsequent motion satisfies the inequality

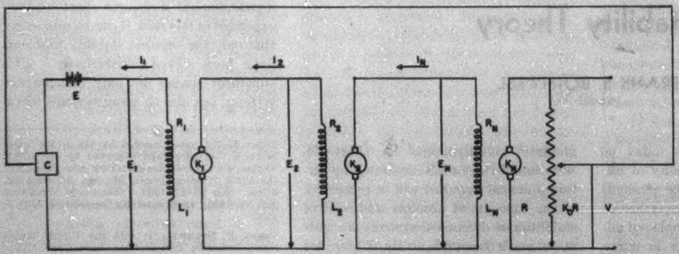


Figure 1. Voltage regulated n-stage d-c generator

$$|x_{i0}(t) - x_i(t)| < \epsilon \qquad (3)$$

If, in addition

$$\lim_{t \to \infty} |x_{i0}(t) - x_i(t)| = 0 \qquad (4)$$

the undisturbed motion is said to be asymptotically stable. If the undisturbed motion $x_{i0}(t)$ is periodic, equation 3 requires the approach of the disturbed motion to a motion of exactly the same period as the period of the undisturbed motion. This last requirement may be relaxed by means of the concept of orbital stability. However, this particular type of stability will not be considered in this paper.

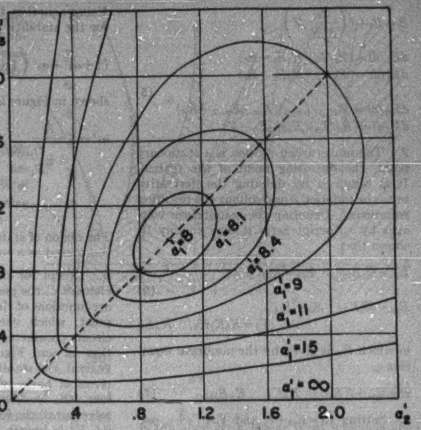


Figure 2. Generalized stability boundary

The First Approximation

The tool which so far has been found most useful (and which falls in the category of the first method) in the investigation of dynamic stability, is the so-called method of first approximation. The procedure is contained in the following six steps.

REPRESENTATION OF THE PHYSICAL SYSTEM BY MEANS OF A SET OF DYNAMICAL EQUATIONS

This first step is often the most difficult. There is no general method of approach, although the resultant dynamical representation is invariably in the form of differential, difference, or integral equations, or, in some cases, a mixed system involving two or three types.

SEARCH FOR THE UNDISTURBED MOTIONS

If the undisturbed motions are stationary points, they are easily found by solving the static equations obtained from the dynamic equations by deleting all derivatives and differences. However, if the undisturbed motions are periodically varying solutions of the dynamical representation, their determination is often most difficult. An approximate means of solution will be provided in the section on the method of equivalent linearization. When the undisturbed motion is nonperiodic and is prescribed by means of certain boundary or other characterizing conditions, there is no general method of approach for nonlinear systems.

PERTURBATIONS

Once the undisturbed motion has been found, all disturbed motions in the neighborhood of the undisturbed motion may be found by means of perturbations. The usual method involves the expansion of the original set of dynamical equations in Taylor series about the undisturbed motion itself. Stability of the set of linear perturbation equations resulting from neglect of the higher order Taylor series terms generally, but not always, guarantees stability of the original dynamical equations. The cases in which the rule breaks down are usually those in which certain linear perturbation terms disappear. Inasmuch as one is usually concerned with the effect on absolute and relative stabilities of system parameter variations, and since the rule breakdown generally occurs only when the system parameters are such that the system itself is on the very edge of instability, such an occurrence requires no further attention. However, there are cases in which the breakdown of the rule occurs for all parameter values, the method of first approximation fails, and one is forced to consider higher order Taylor series terms.

CHARACTERISTIC EQUATION

If the original dynamical equations are autonomous, that is, if the dependence on time is not explicit, and if the undisturbed motion is a stationary point, the equations of first approximation are linear and possess constant coefficients. Stability is consequently determined by the roots of the characteristic equation of the first approximation. If the original system is described entirely by differential equations or entirely by difference equations, the characteristic equation is either an algebraic equation or can be made so by a logarithmic transformation. However, if the system is represented by a mixed system of differential-difference equations, the characteristic equation is transcendental and presents certain difficulties of solution.

STABILITY CRITERIA

The occurrence of one or more roots of the characteristic equation with positive real part is the necessary and sufficient condition for instability of the original system. Whether or not this situation occurs can be determined by means of either the numerical method of Nyquist[3] or the analytical method of Routh[4] and Hurwitz.[5] The choice of method depends on the complexity of the characteristic equation and on whether one is concerned with parameter variations.[6] Both the Routh method and the Nyquist method were originally derived for the treatment of absolute stability. However each may be modified to yield certain information concerning the actual location of the roots.[6–9]

STABILITY BOUNDARIES

The final step in any stability analysis is the presentation of the absolute and relative stability boundaries as functions of basic system parameters. A few examples will now be given.

Voltage Regulation[10]

Let us follow through the six steps for the n-stage d-c generator, shown in Figure 1, which employs a 3-terminal nonlinear regulator C, such as the step resistor, the resistance value of which varies in steps with the control voltage V.

1. Denoting the voltage function of the regulator C by $f(I_1, V)$, the dynamical equations are

$$E=E_1+f\left(\frac{E_2}{K_1R_1},V\right)$$

$$\begin{aligned}dE_2/dt&=(R_1/L_1)(K_1E_1-E_2)\\ dE_3/dt&=(R_2/L_2)(K_2E_2-E_3)\\ &\dots\\ dE_n/dt&=(R_{n-1}/L_{n-1})(K_{n-1}E_{n-1}-E_n)\\ dV/dt&=(R_n/L_n)(K_0K_nE_n-V)\end{aligned}\qquad(5)$$

2. The undisturbed motion is a stationary point, the operating point of the system. It is obtained by deleting the derivative terms in equation 5 and solving the resulting equations. Denoting the equilibrium voltages by subscript zero, it may readily be shown that

$$\begin{aligned}&E_{20}=K_1E_{10}\qquad E_{30}=K_1K_2E_{10}\\ &\dots\\ &E_{n0}=K_1K_2\dots K_{n-1}E_{10}\\ &\qquad V_0=K_0K_1K_2\dots K_nE_{10}\end{aligned}\qquad(6)$$

in which E_{10} is given by the nonlinear equation

$$E=E_{10}+f(E_{10}/R_1,K_0K_1\dots K_nE_{10})\qquad(7)$$

3. Setting $E_i=E_{i0}+e_i$ and $V=V_0+v$, one finds the set of linear perturbation equations

$$\begin{aligned}&e_1+(f_{I_1}/K_1R_1)e_2+f_Vv=0\\ &de_2/dt+(R_1/L_1)e_2-K_1(R_1/L_1)e_1=0\\ &\dots\\ &de_n/dt+(R_{n-1}/L_{n-1})e_n-K_{n-1}\times\\ &\qquad(R_{n-1}/L_{n-1})e_{n-1}=0\\ &dv/dt+(R_n/L_n)v-K_0K_n(R_n/L_n)e_n=0\end{aligned}\qquad(8)$$

in which f_{I_1} and f_V denote the partial derivatives of f with respect to I_1 and V, evaluated at the equilibrium point.

4. The characteristic equation 8 is obtained by equating to zero the determinant of the coefficients of the unknown voltages e_i and v, after replacing the operation of differentiation by the symbol p. It is

$$\left(p+\frac{R_n}{L_n}\right)\left(p+\frac{R_{n-1}}{L_{n-1}}\right)\dots\left(p+\frac{R_2}{L_2}\right)\times\left[p+\frac{R_1}{L_1}\left(1+\frac{f_{I_1}}{R_1}\right)\right]+K\frac{R_1R_2\dots R_n}{L_1L_2\dots L_n}f_V=0\qquad(9)$$

in which $K=K_0K_1\dots K_n$

5. The Hurwitz determinant is useful in this case. For a 3-stage generator, equation 9 is the cubic

$$a_0p^3+a_1p^2+a_2p+a_3=0\qquad(10)$$

in which

$$\begin{aligned}a_0&=1\\ a_1&=(R_3/L_3)+(R_2/L_2)+(R_1/L_1)\left(1+\frac{f_{I_1}}{R_1}\right)\\ a_2&=\frac{R_3R_2}{L_3L_2}+\left(\frac{R_3}{L_3}+\frac{R_2}{L_2}\right)\frac{R_1}{L_1}\left(1+\frac{f_{I_1}}{R_1}\right)\\ a_3&=\frac{R_3R_2R_1}{L_3L_2L_1}\left(1+\frac{f_{I_1}}{R_1}+Kf_V\right)\end{aligned}\qquad(11)$$

The stability boundary is obtained by equating to zero the penultimate Hurwitz determinant

$$\begin{vmatrix}a_1 & a_0\\ a_3 & a_2\end{vmatrix}$$

Using equation 11, one finds an expression for the stability boundary

$$(1+\alpha_2'+\alpha_3')\left(1+\frac{1}{\alpha_2'}+\frac{1}{\alpha_3'}\right)=1+\alpha_1'\qquad(12)$$

shown in Figure 2, in which

$$\alpha_1'=K\frac{f_V}{1+\frac{f_{I_1}}{R_1}},\ \alpha_2'=\frac{R_2}{L_2}\,\frac{L_1}{R_1+f_{I_1}},$$

$$\alpha_3'=\frac{R_3}{L_3}\,\frac{L_1}{R_1+f_{I_1}}\qquad(13)$$

The region of stability lies below the surface considered as a contour map.

6. Except for the case of a linear control element C, the partial derivatives f_{I_1} and f_V are functions of the location of the operating point, which itself is determined by the circuit parameter values. Therefore equation 12 and Figure 2 do not express, in general, the stability boundary explicitly in terms of basic system parameters, and, in fact, the problem cannot be completely solved until the form of the control function $f(I_1,V)$ is known. A common control element for high-voltage generators is the step resistor. The essential features of its stability in connection with Figure 1 may be observed by assuming that the resistance varies linearly and continuously with control voltage when the latter exceeds a critical value, V_c. Thus

$$\begin{aligned}R_c&=0 &&\text{if } V<V_c\\ &=K_c\frac{R_1}{V_c}(V-V_c) &&\text{if } V>V_c\end{aligned}$$

so that

$$\begin{aligned}f(I_1,V)&=0 &&\text{if } V<V_c\\ &=K_c\frac{R_1}{V_c}(V-V_c)I_1 &&\text{if } V>V_c\end{aligned}$$

and the stability boundary takes the form of Figure 3, or

$$\left[\frac{1}{2}(\sqrt{1+\alpha_1''}-1)+\alpha_2''+\alpha_3''\right]\times\left[1+\frac{1}{2}(\sqrt{1+\alpha_1''}-1)\times\left(\frac{1}{\alpha_2''}+\frac{1}{\alpha_3''}\right)\right]=\sqrt{1+\alpha_1''}\qquad(14)$$

in which

$$\alpha_1''=4\frac{K_c}{(K_c-1)^2}K_0K_1\dots K_n\frac{E}{V_c}$$

$$\alpha_2''=\frac{1}{K_c-1}\,\frac{R_2L_1}{L_2R_1}\qquad\alpha_3''=\frac{1}{K_c-1}\,\frac{R_3L_1}{L_3R_1}\qquad(15)$$

In Figure 3, the region of instability lies inside the mound enclosed by the coordinate plane $\alpha_1''=0$ and the contour surface.

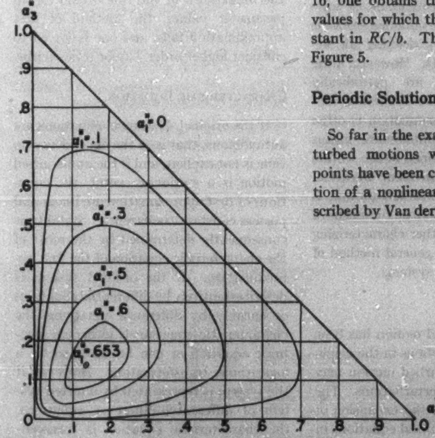


Figure 3. Stability boundary for a controlled resistance as control element

Shunt Feedback Amplifier[6]

An excellent example of the concept of relative stability is afforded by the treatment of the circuit of Figure 4. For the case in which the admittance Y is composed of a parallel combination of resistance and capacitance R and C, one system time constant is always equal to RC. Denoting by RC/b, where $-\infty<b<1$, the value of the maximum time constant of the system, the region of relative stability is given by the inequality

$$\begin{aligned}\alpha_3&<\frac{1}{\alpha_1-\alpha_2} &&\text{for } \alpha_2<\frac{2}{3}\alpha_1\\ &<\frac{2(2+\alpha_1\alpha_2)^2}{2\alpha_1+\alpha_2} &&\text{for } \alpha_2>\frac{2}{3}\alpha_1\end{aligned}\qquad(16)$$

in which

$$\alpha_1=K_2,\ \alpha_2=\frac{K_1RS_2}{1-b},\ \alpha_3=\frac{R^2S_1S_2}{(1-b)^2}$$

If equality signs are inserted in equation 16, one obtains the locus of parameter values for which the maximum time constant in RC/b. The results are shown in Figure 5.

Periodic Solutions

So far in the examples only the undisturbed motions which were stationary points have been considered. The operation of a nonlinear oscillator is often described by Van der Pol's equation

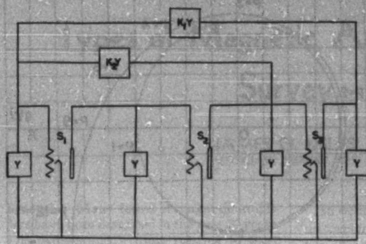


Figure 4 (left). Illustrative circuit—shunt feedback amplifier

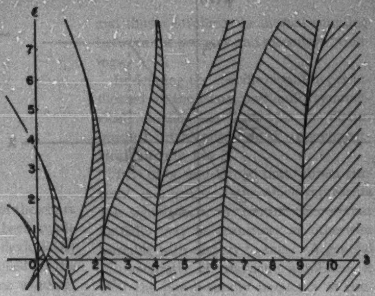


Figure 6 (left). Stable and unstable regions of the Mathieu equation

$$\frac{d^2x}{dt^2}+x+\mu(x^2-1)\frac{dx}{dt}=0 \tag{17}$$

For μ small, one undisturbed motion is given approximately by

$$x_0(t)=2\sin t \tag{18}$$

The equations of first approximation are obtained by the substitution of

$$x(t)=x_0(t)+\delta(t)$$

together with equation 18, into equation 17 and omitting all but linear terms in δ. The result is the linear differential equation with periodic coefficients

$$\frac{d^2\delta}{dt^2}+\mu(4\sin^2 t-1)\frac{d\delta}{dt}+(1+8\sin t\cos t)\delta=0 \tag{19}$$

Here the methods of Nyquist and Routh are of no use, and the general method, as outlined by Stoker,[11] for example, stops far short of the actual solution for stability boundaries. In fact, the boundaries can be obtained only after one has evaluated a certain infinite order determinant called a Hill determinant.

For the very simple case of the Mathieu equation

$$\frac{d^2x}{dt^2}+(\delta+\epsilon\cos t)x=0 \tag{20}$$

which represents a variety of physical systems,[12] the stability boundary has been completely determined and is shown in Figure 6. Even in this simple case the boundary is a very complicated function of the parameters δ and ϵ.

Equivalent Linearization

A very powerful tool for the investigation of stability of periodic motion is based on the so-called method of equivalent linearization as formulated by Kryloff and Bogoliuboff,[13] and apparently first linked with the Nyquist criterion by Kochenburger[14] in his treatment of the relay servomechanism. The method involves the systematic replacement of all nonlinear elements in the system by quasi-linear elements which have the property of preservation of waveform but nonlinear magnification. For example, in the Van der Pol equation 17, the nonlinear term

$$(x^2-1)\frac{dx}{dt} \tag{21}$$

is replaced by a linear term which is the result of considering only the fundamental term in the response of equation 21 to the excitation $x=a\sin\omega t$. But the fundamental term is $(a^2/4-1)a\omega\cos\omega t$, so that the equivalent linear term is $(a^2/4-1)(dx/dt)$, and equation 17 is replaced by

$$\frac{d^2x}{dt^2}+x+\mu\left(\frac{a^2}{4}-1\right)\frac{dx}{dt}=0 \tag{22}$$

Thus oscillations which have amplitude greater than $a=2$ are obviously damped, while oscillations of magnitude less than $a=2$ obviously are characterized by negative damping. Therefore, the solution $x=2\sin t$ is stable.

Inasmuch as the disturbed motion (and for that fact even the undisturbed motion) is not exactly sinusoidal, there is some doubt as to what is meant by the amplitude of motion. In fact, the whole procedure rests on rather flimsy mathematical ground. However, in systems which are essentially low-pass filters, such as servomechanisms, wherein harmonic distortion is low, the procedure has been very successful, and this is its great redeeming feature.

A remarkable example of the use of equivalent linearization occurs in the field of economics. A model of the business cycle has been hypothesized by Goodwin[15] in the form of the autonomous nonlinear mixed differential-difference equation

$$\gamma y(t+\theta)+(1-\alpha)y(t+\theta)=\phi[y(t)] \tag{23}$$

in which $y(t)$ denotes present national income, and the induced investment function $\phi(x)$ is the unsymmetrical saturation curve of Figure 7. If c denotes the amplitude of the quasi-linear oscillation of $y(t)$, the characteristic equation of the

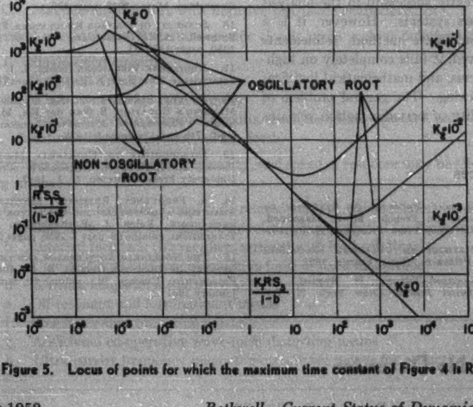


Figure 5. Locus of points for which the maximum time constant of Figure 4 is RC/b

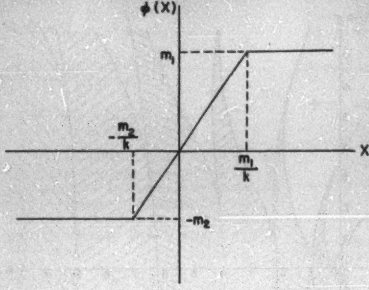


Figure 7 (left). Induced investment function

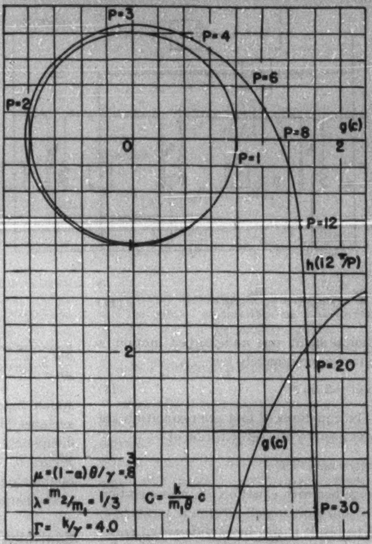


Figure 8 (right). Nyquist plot of quasi-linear representation of Goodwin's model of the business cycle

corresponding linearized equation is found to be of the form

$$h(s)=g(c) \tag{24}$$

The Nyquist plot of each side of equation 24 is shown in Figure 8. Here c is plotted as a function of $g(c)$ so that the intersection of the spiral $h(i\omega)$ ($\omega=2\pi/p$, where p is the period of the assumed oscillation) with the $g(c)$ axis determines an infinite number of admissible values of period p. That is, it is possible that the system oscillates in any one of an infinite number of modes. The admissible values of p are approximately 8 years, 1 year, 1/2 year, 1/3 year, and so forth.

The question of stability is decided by the following reasoning. For values of c for which the spiral encloses the point $g(c)$, the motion is stable, c will decrease, and $g(c)$ will increase. For values of c for which the spiral does not enclose the point $g(c)$, the motion is unstable, c will increase, and $g(c)$ will decrease. Obviously each of the infinite number of modes is stable. It is also apparent that a relatively small disturbance of a high order mode is sufficient to cause oscillation in a different mode, a result which is not available in the analysis of infinitesimal disturbances. That these modes do exist and are stable, has been demonstrated experimentally by means of analogue equipment.

Conclusions

Of the three types of stability, orbital, temporal, and asymptotic, the last has received the most attention. Almost all the successful stability analyses have employed the first of Liapounoff's two methods, and have been concerned mainly with systems subject only to infinitesimal disturbances.

Undisturbed motions which are stationary points almost always are treated successfully by a sequence of first order perturbation and a stability criterion of Nyquist or Routh type. Relative stability as well as absolute stability may be treated by this method.

Undisturbed motions which are periodic are currently treated by two methods. The perturbation method invariably produces a linear equation with periodic coefficients, the stability of which may be investigated by Hill's method. However the general method stops far short of a solution, and only the simplest cases have been solved. Equivalent linearization has been currently successful in the analysis of low-pass systems. However, it is a very approximate method, refinements are involved, it fails completely on high-pass systems, and mathematical justification is lacking. The general problem of the stability of periodic motion remains unsolved.

References

1. The General Problem of Stability of Motion (book) (in French), **M. A. Liapounoff.** Princeton University Press, Princeton, N. J., 1947.

2. New Methods in Celestial Mechanics (book), **H. Poincare.** Paris, France, 1892.

3. Regeneration Theory, **H. Nyquist.** *Bell System Technical Journal* (New York, N. Y.), volume 11, January 1932, pages 126–47.

4. Stability of a Given State of Motion, **E. J. Routh.** Adams Prize Essay. Macmillan, London, England, 1877.

5. On the Condition for which an Equation Possesses only Roots with Negative Real Parts, **A. Hurwitz.** *Mathematische Annalen* (Leipzig, Germany), volume 46, 1895, pages 273–84.

6. Nyquist Diagrams and the Routh-Hurwitz Stability Criterion, **F. E. Bothwell.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 38, number 11, November 1950.

7. A Generalization of Nyquist's Stability Criteria, **A. Vazsonyi.** *Journal of Applied Physics*, American Institute of Physics (New York, N. Y.), volume 20, September 1949, pages 863–67.

8. The Location of Zeros of Polynomials with Complex Coefficients, **E. Frank.** *Transactions*, American Mathematical Society (New York, N. Y.), volume 52, October 1946, pages 890–98.

9. Foundations of Economic Analysis (book), **P. A. Samuelson.** Harvard University Press, Cambridge, Mass., 1948, page 435.

10. Stability of Voltage Regulators, **Frank E. Bothwell.** *AIEE Transactions*, volume 69, part II, 1950, pages 1430–33.

11. Nonlinear Vibrations (book), **J. J. Stoker.** Interscience, New York, N. Y., 1950, pages 189–222.

12. On the Stability of the Solutions of Mathieu's Equation, **B. Van der Pol, M. J. O. Strutt.** *Philosophical Magazine*, (London, England), 7th series, volume 5, 1928.

13. Introduction to Nonlinear Mechanics (book), **N. Kryloff, N. Bogoliuboff.** Princeton University Press, Princeton, N. J., 1943.

14. A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms, **Ralph J. Kochenburger.** *AIEE Transactions*, volume 69, part I, 1950, pages 270–83.

15. The Nonlinear Accelerator and the Persistence of Business Cycles, **R. M. Goodwin.** *Econometrica* (Chicago, Ill.), volume 19, number 1, January 1951.

No Discussion

Types of Magnetic Amplifiers—Survey

JAMES G. MILES
ASSOCIATE MEMBER AIEE

Synopsis: Many forms of magnetic amplifiers have been described in the past 50 years. The patents and writings relating to these devices have been examined to determine what types of magnetic amplifiers exist and what properties these devices have in common. The objects of this paper are: (1) to describe the various types of magnetic amplifiers, and (2) to present a method by which the basic principles relating to these devices may be codified according to an orderly system.

Magnetic Amplifier Chart

A chart which is basic to a scheme for classifying the various types of magnetic amplifiers is given below. Any known magnetic amplifier may be represented by a combination of the key letters and numbers listed in the chart. Each of the letters and numbers represents information relating to the magnetic amplifier class, circuit type, and special means for affecting the characteristics of the particular amplifier to be described.

In defining classes A and B of the chart an arbitrary distinction has been made between inductive-impedance devices (A) in which the a-c power is applied to the core via the load or reactance winding, and (B) in which the a-c power is applied to the core via a separate primary winding. According to this distinction, therefore, key A refers to saturable reactors controlled by signal magnetomotive force and key B refers to saturable transformers controlled by signal magnetomotive force.

Two classes of signal-magnetomotive-force-controlled saturable transformers appear. The first, key B′, relates to a class of saturable transformers in which control is effected by varying the incremental permeability or by otherwise distorting the flux-time wave form in the core. The second class, key B″, relates to the class of transformers in which control is effected by varying the reluctance of a magnetic path which shunts the transformer leg linking the secondary winding and thus controlling the mutual coupling between the primary and secondary windings.

In defining the specific circuit types indicated on the chart, push-pull (symmetrical) circuits, key 1, are to be distinguished from single-sided (unsymmetrical) circuits, key 2.

It is to be understood that "direct current" is a generic term which represents the signal, bias, or control effects that actually may be transitory or otherwise variable with respect to time. A general characteristic of magnetic amplifiers is that the power-supply frequency is considerably higher than any component of input-signal frequency.

To simplify the classification only single-stage magnetic amplifiers are considered. Cascaded arrangements of stages are not included, as such, in this classification, but considerations are stated which must be taken into account in the design of single magnetic-amplifier stages which are to be used in cascade with other stages.

Classification Chart for Magnetic Amplifiers

General Classes

- A. Saturable reactors controlled by signal magnetomotive force
- B. Saturable transformers controlled by signal magnetomotive force
    - B′. Control by varying incremental permeability or by distorting wave form
    - B″. Control of mutual inductance by varying the reluctance of a saturable shunt
- C. Magnetic modulators with negative resistance characteristics

General classes A and B contain specific circuit types such as polarized and nonpolarized magnetic amplifiers comprising:

1. Single-sided circuits
2. Push-pull circuits
3. Wheatstone-bridge-type circuits
4. Special magnetic amplifier circuits (such as cross-connected pairs of magnetic amplifiers, magnetic amplifiers in combination with transformers, and so forth)

Amplifier sensitivity, stability, linearity, and speed of response may be affected by incorporating any one or more of the following means:

- a. External feedback (amplitude-related and/or phase-related external feedback means including impedance coupled means terminating in a single-control winding or special feedback winding)
- b. Internal feedback (by rectifiers seriesed with load windings)
- c. Unidirectional bias magnetomotive force of fixed magnitude
- d. A-c bias magnetomotive force at fundamental or harmonic of power supply (carrier) frequency
- e. Off-resonant and ferroresonant means
- f. Frequency-discriminating means
- g. Additional co-operative wave-form distorting means
- h. Ultra-carrier frequency magnetomotive force means for reduction of hysteresis effects

---

Paper 52-216, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted December 4, 1951; made available for printing May 14, 1952.

James G. Miles is with Engineering Research Associates, Inc., St. Paul, Minn.

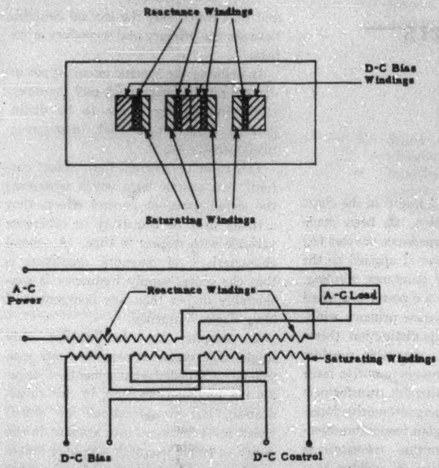

Reactance Windings
D-C Bias Windings
Saturating Windings
A-C Power
Reactance Windings
A-C Load
Saturating Windings
D-C Bias
D-C Control

Figure 1. Single-sided d-c saturable reactor without feedback

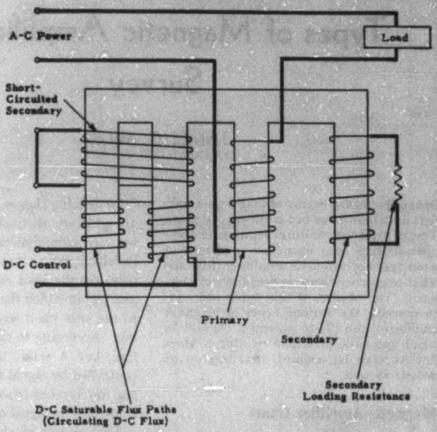

A-C Power
Load
Short-Circuited Secondary
D-C Control
Primary
Secondary
Secondary Loading Resistance
D-C Saturable Flux Paths (Circulating D-C Flux)

Figure 3. Variable mutual-inductance type d-c saturable transformer

Examples of Types of Magnetic Amplifiers

Figures 1 through 20 show typical magnetic amplifiers which may be classified according to the chart. At least one example of a circuit relating to each key is shown in the figures. The examples have been chosen not on the basis of dates of invention, but because they illustrate particularly well the circuit types described.

Figure 1 (class *A*, *1*, *c*) is a schematic of the first known d-c controlled saturable reactor for which gain was claimed. E. F. W. Alexanderson[1] described this device in United States Patent *1,206,643* (1916), and as a basic device it has been used widely. Alexanderson's device comprises two variable-reactance windings as load-current control means, the reactance windings being placed in series between the a-c load and the a-c power supply. The total impedance of the circuit is controlled according to the reactance of the reactance windings. Control is accomplished in a well-known manner based on the change of incremental permeability of the core according to the amount of d-c saturating magnetomotive force applied to the core. The saturating windings are connected so that opposing a-c electromotive forces are induced in them by the reactance windings; the result is that zero alternating voltage of fundamental power-supply frequency appears across the terminals of the d-c control circuit. Alexanderson provided separate d-c bias windings for the purpose of introducing a d-c bias magnetomotive force to locate an optimum operating point on the magnetic characteristics.

Figure 2 is an example of a d-c controlled saturable transformer of the variable incremental-permeability type (class *B′*, *1*). This circuit is due to A. E. Bowen,[2] United States Patent *2,230,558* (1941). Here the primary and secondary windings are wound on the outside legs of a 3-legged core in such directions as to be mutually coupled and so as to pass no flux of power-supply fundamental frequency through the center leg of the core. Control is effected by imposing, via the saturating winding, a condition of d-c premagnetization. If a core is premagnetized by a d-c component of magnetomotive force, a given a-c magnetomotive force introduced by the primary winding will effect less change in flux than if the core were not premagnetized. Secondary voltage therefore may be controlled in a core by introducing a controlled amount

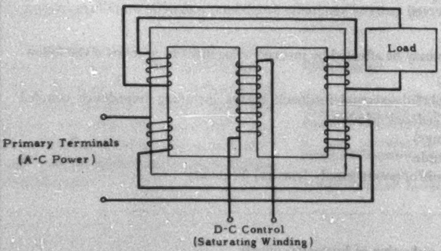

Load
Primary Terminals (A-C Power)
D-C Control (Saturating Winding)

Figure 2. Variable-permeability type d-c saturable transformer

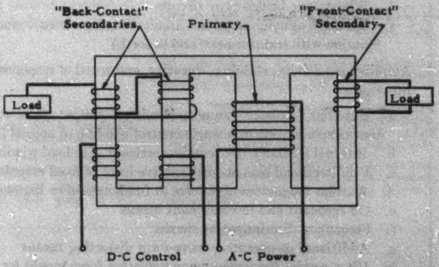
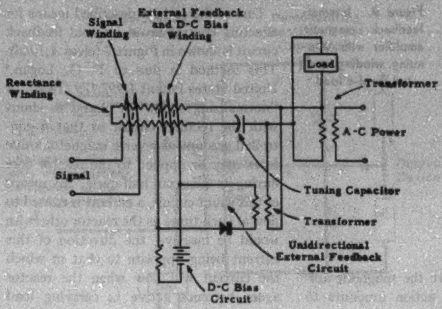

"Back-Contact" Secondaries
Primary
"Front-Contact" Secondary
Load
Load
D-C Control
A-C Power

Figure 4. Variable mutual-inductance type d-c saturable transformer


Signal Winding
External Feedback and D-C Bias Winding
Reactance Winding
Load
Transformer
A-C Power
Signal
Tuning Capacitor
Transformer
Unidirectional External Feedback Circuit
D-C Bias Circuit

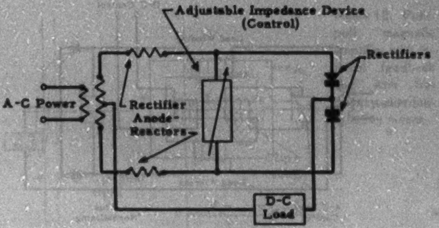

Adjustable Impedance Device (Control)
Rectifiers
A-C Power
Rectifier Anode-Reactors
D-C Load

Figure 5 (right). Single-sided d-c saturable reactor with combined external feedback

Figure 7 (above). Flux-normalizing method of control for internal-feedback magnetic amplifier

of d-c premagnetization so as to effect a control of the a-c flux variation in the magnetic circuit linking the secondary winding with the primary winding. The amplitude of the a-c flux which mutually couples the primary and secondary is controlled by the degree of d-c premagnetization.

A d-c saturable-shunt type of variable-mutual-inductance transformer is shown in Figure 3. This circuit, due to E. R. Stoekle,[3] United States Patent *1,376,978* (1921), (class *B″*, *1*), shows a 4-legged transformer in which the flux that couples the secondary with the primary may be shunted, to a controlled degree, by a variable reluctance shunt. The mutual flux that links the secondary with the primary winding thus is decreased and controlled effectively. Control is effected in this typical circuit by means of saturating windings arranged to cause "circulating" d-c flux in the two saturable shunt legs. In this manner, the reluctance of the shunt legs to the passage of a-c flux is controlled effectively. In Figure 4, the transformer primary winding is inserted as a controlled reactance in series between the load and a-c power source. The reactance of the primary winding in this circuit is controlled according to the combined reflected impedances from the two secondary windings with which it is immediately coupled, the reflected impedances being controlled according to the relative mutual coupling between the primary and secondary windings.

Another example of the d-c saturable shunt-type variable-mutual-inductance transformer is that shown in Figure 4. C. M. Hines[4] in United States Patent *2,215,820* (1940) described a saturable transformer (class *B″*, *1*) similar to that shown in Figure 3 but having another set of secondary windings placed on the d-c saturable legs as shown. These "back contact" secondary windings are arranged to be coupled effectively with the primary windings when the control legs are desaturated. On the other hand, if the d-c control legs are saturated, a minimum of primary flux will link the back contact secondary windings and a maximum of primary flux will link the "front contact" secondary winding. This circuit, with the primary winding excited by the a-c power supply and with the secondary windings exciting separate loads, is, in effect, a single-pole double-throw a-c power switch.

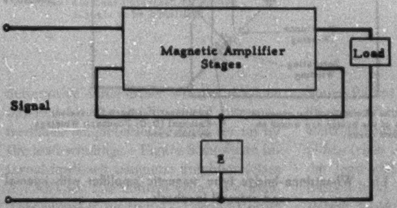
Means are abundant in the magnetic-amplifier art by which either positive or negative feedback may be introduced to affect gain, stability, linearity, and speed of response of magnetic amplifiers. Feedback means may be divided into two principal classes: external feedback and internal feedback. These classes are described in the following paragraphs.

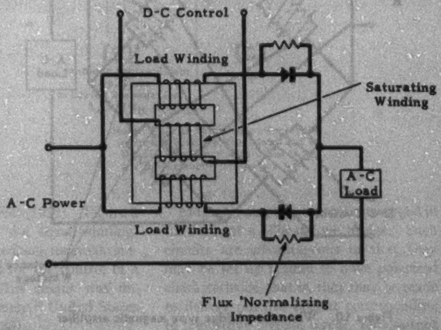
In external feedback circuits, a measure of output voltage and/or current is reintroduced as a feedback magnetomotive force via an external circuit which may terminate either in the control winding or in a special feedback winding. The amount of external feedback may be modified by adjusting either the amplitude or the phase of the feed-back signal with respect to the input or control signal.

One of the earliest examples of the application of amplitude-related external feedback to a d-c saturable-reactor magnetic amplifier is shown in Figure 5 (class *A,1,a,c,f*). Here E. F. W. Alexanderson[5]

Figure 6 (below). Impedance-coupled external feedback means

Figure 8 (right). Internal-feedback magnetic amplifier with saturating winding control for a-c load


Magnetic Amplifier Stages
Load
Signal
Z


D-C Control
Load Winding
Saturating Winding
A-C Power
A-C Load
Load Winding
Flux Normalizing Impedance

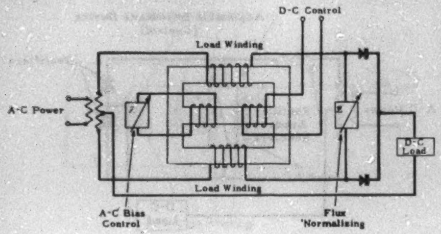


Figure 9. Internal-feedback magnetic amplifier with saturating winding control for d-c load

in United States Patent *1,328,797* (1920) showed a method for obtaining load-current control by placing a controlled reactance winding in parallel with the load and across the secondary terminals of an a-c supply transformer. A measure of magnetic-amplifier output potential (also load potential) is applied via a second transformer and rectifier circuit to an external feedback winding in such manner as to aid the effect of the control magnetomotive force. D-c bias is introduced via a parallel circuit using the same winding as used for external feedback, as shown.

An alternative means for introducing amplitude-related external feedback is shown in Figure 6 which illustrates, in block diagram form, an impedance-coupled external feedback means (class -, *1,a*) described by A. S. FitzGerald[6] in United States Patent *2,464,639* (1949). A measure of load current creates a potential drop across the impedance *Z*. The impedance is so connected that the potential drop across it will aid (or oppose) the control signal applied at the magnetic-amplifier input. This action amounts to feedback.

A phase-related external feedback circuit is described by J. Slepian[7] in United States Patent *1,645,302* (1927), not illustrated here. In this patent a series-resonant inductance - capacitance phase shifter in the control circuit fixed the phase of the control signal with respect to the feed-back signal.

The so-called "internal feedback" means is the second important means for effecting feedback. In order to achieve internal feedback, rectifiers are inserted in series between the load windings and the load circuit. Since only unidirectional current impulses can flow in the load windings of such circuits, unidirectional pulsating flux components are created in the reactor cores in magnitude according to the magnitude of the load current. Control of such devices is effected in one of several ways to be discussed in the following paragraphs.

One of the earliest described means for effecting control in an internal feedback circuit is shown in Figure 7 (class *A,1,b,d*). This method is due to F. G. Logan,[8] United States Patent *1,997,179* (1935). A "normalizing shunt" is placed in shunt with the rectifier circuit so that a controlled amount of reverse magnetomotive force may be applied to the reactor core during its quiescent half-cycle. By means of this shunt circuit, a current is caused to flow at such times as the reactor otherwise would be inactive, the direction of this current being opposite to that in which the current will flow when the reactor again becomes active in carrying load current. The flux and energy conditions thus tend to establish, to a degree determined by the adjustment of the normalizing impedance, the conditions which would have existed if the reactor had not been subjected to an inactive condition. According to the value of the normalizing impedance, therefore, a controlled amount of energy and current is delivered by the load circuit. The changes in normalizing-impedance current required to effect control are claimed to be very slight compared to the changes in load current which may be obtained.

Another and more widely used means of control for internal-feedback magnetic amplifiers is shown in Figures 8 and 9. This means of magnetic-amplifier control employs a d-c saturating winding to presaturate the core by an amount adjusted by the control current. The control winding links the magnetic core arrangement in such manner that the d-c control mag-

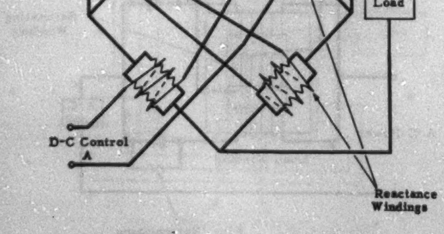
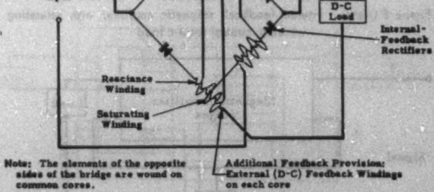


Figure 10. Wheatstone-bridge type magnetic amplifier



Note: The elements of the opposite sides of the bridge are wound on common cores.

Additional Feedback Provision: External (D-C) Feedback Windings on each core

Figure 11. Wheatstone-bridge type magnetic amplifier with internal and external feedback

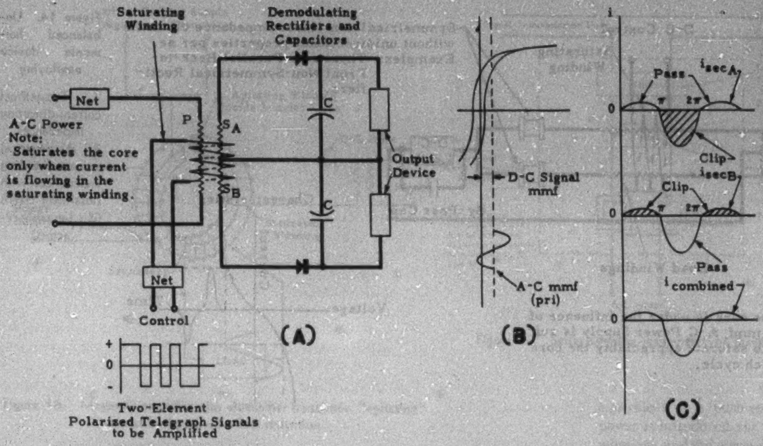

Saturating Winding
Demodulating Rectifiers and Capacitors
Net
A-C Power
Note: Saturates the core only when current is flowing in the saturating winding.
P
S_A
S_B
C
C
Output Device
Net
Control
(A)
+
0
-
Two-Element Polarized Telegraph Signals to be Amplified
D-C Signal mmf
A-C mmf (pri)
(B)
i
0
Pass
i_{secA}
π
2π
Clip
i_{secB}
Clip
0
π
2π
Pass
$i_{combined}$
0
(C)

Figure 12. Push-pull magnetic amplifier with internal feedback and with frequency-discriminating means

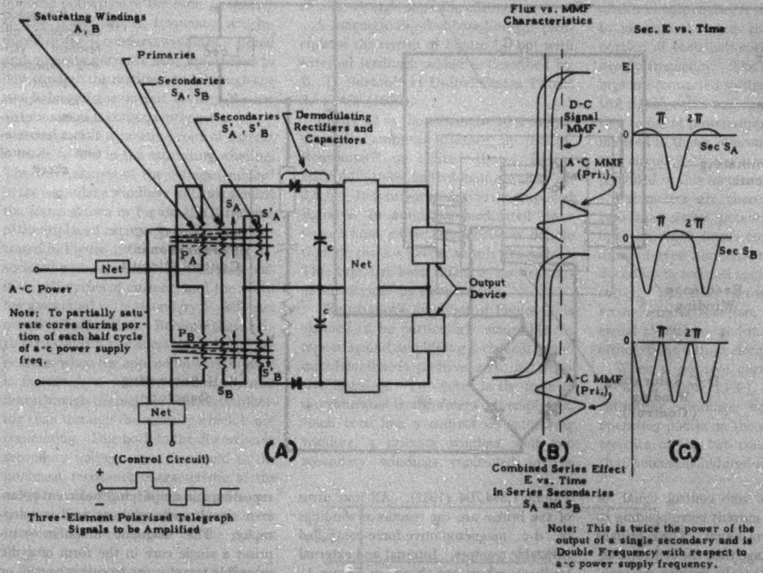

Saturating Windings A, B
Primaries
Secondaries S_A, S_B
Secondaries S'_A, S'_B
Demodulating Rectifiers and Capacitors
S_A
S'_A
P_A
Net
A-C Power
Note: To partially saturate cores during portion of each half cycle of a-c power supply freq.
c
c
Net
Output Device
P_B
S'_B
S_B
Net
(Control Circuit)
+
0
−
Three-Element Polarized Telegraph Signals to be Amplified
(A)
Flux vs. MMF Characteristics
D-C Signal MMF.
A-C MMF (Pri.)
A-C MMF (Pri.)
(B)
Combined Series Effect E vs. Time In Series Secondaries S_A and S_B
Sec. E vs. Time
E
0
π
2π
Sec S_A
0
π
2π
Sec S_B
0
π
2π
(C)
Note: This is twice the power of the output of a single secondary and is Double Frequency with respect to a-c power supply frequency.

Figure 13. "Second - harmonic type" push-pull magnetic amplifier with internal feedback and with frequency - discriminating means

netomotive force either will be aided or opposed by the unidirectional internal-feedback magnetomotive force set up by the load windings. Figure 8 shows an internal-feedback magnetic amplifier (class *A,1,b,d*) for the control of an a-c load. This circuit is due to F. G. Logan,[9] United States Patent *2,126,790* (1938). Figure 9 illustrates a similar d-c signal-winding-controlled internal-feedback magnetic amplifier (class *A,1,b,d*) for the control of a d-c load circuit. This circuit was described by S. M. Hanely,[10] United States Patent *2,144,290* (1939).

Magnetic amplifiers may be arranged in the form of a Wheatstone bridge. Such circuits are advantageous in that they may be set up readily to have polarized characteristics, that is, they may be made to deliver positive output corresponding to a positive control signal, zero output

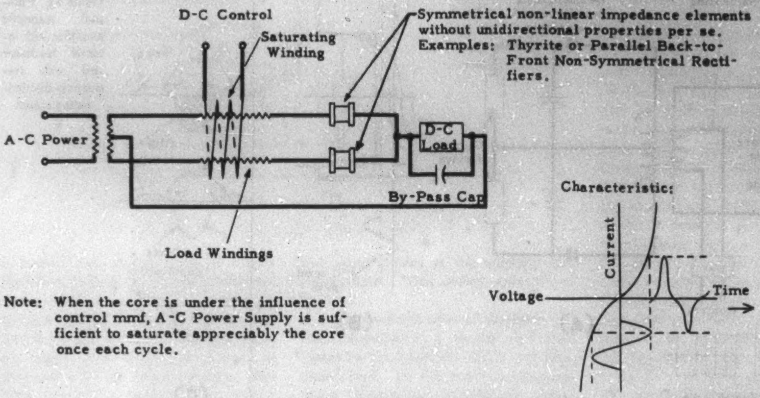

D-C Control
Saturating Winding
Symmetrical non-linear impedance elements without unidirectional properties per se.
Examples: Thyrite or Parallel Back-to-Front Non-Symmetrical Rectifiers.
A-C Power
D-C Load
By-Pass Cap
Load Windings
Note: When the core is under the influence of control mmf, A-C Power Supply is sufficient to saturate appreciably the core once each cycle.
Characteristic:
Current
Voltage
Time

Figure 14. Unbalanced harmonic device employing

1. Symmetrical current-distorting impedances
2. Asymmetrical current-distorting impedances with polarizing means (d-c saturable reactors)

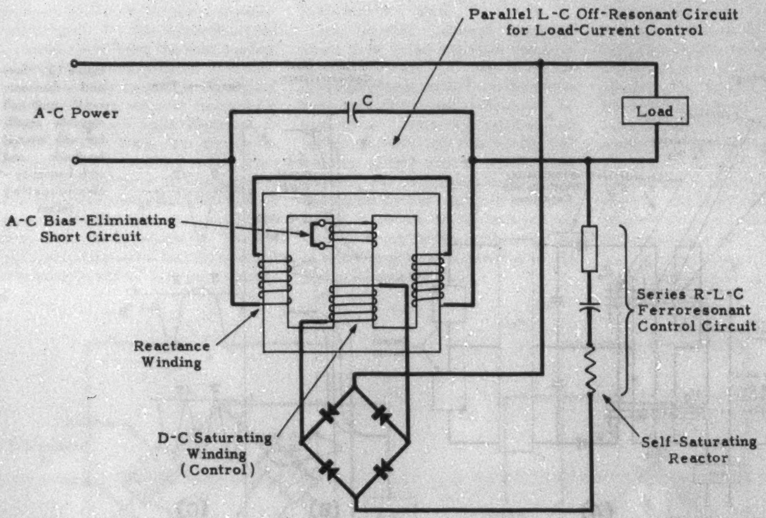

Parallel L-C Off-Resonant Circuit for Load-Current Control
A-C Power
C
Load
A-C Bias-Eliminating Short Circuit
Reactance Winding
D-C Saturating Winding (Control)
Series R-L-C Ferroresonant Control Circuit
Self-Saturating Reactor

Figure 15. Regulating apparatus with

1. Inductance-capacitance off-resonant variable - impedance circuit
2. Resistance-inductance - capacitance ferroresonant control circuit

corresponding to zero control signal, or negative output current corresponding to a negative control signal. These bridges have the advantage that the control, output, and power-supply circuits are practically isolated one from the other.

Many Wheatstone-bridge circuits have been described in which one or more arms of the bridge comprise the reactance windings of magnetic amplifiers. A specific example is the class *A,3* circuit shown in Figure 10, this circuit being due to L. W. Thompson,[11] United States Patent *1,844,704* (1932). All four arms of the bridge are the reactance windings of d-c magnetomotive-force-controlled saturable reactors. Internal and external feedback means are shown in Figure 11 as applied to the Wheatstone-bridge type magnetic amplifier due to Uno Lamm,[12] United States Patent *2,403,891* (1946) (class *A,3,a,b*).

A class *B′,2,b,f,g* device is disclosed by E. M. Boardman[13] in United States Patent *2,108,642* (1938). This device, schematically shown in Figure 12(A), is for repeating and amplifying 2-element polarized signals such as those used in telegraphy. The magnetic amplifier comprises a single core in the form of a d-c saturable transformer having a control or saturating winding, a primary winding, and two secondary windings connected in a circuit which is the equivalent of a push-pull system. The amount of a-c power magnetomotive force introduced via the primary winding normally is not sufficient to saturate the core unless control magnetomotive force also exists.

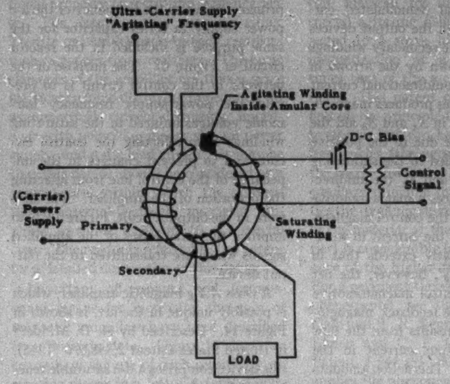

Ultra-Carrier Supply
"Agitating" Frequency
Agitating Winding
Inside Annular Core
D-C Bias
Control
Signal
(Carrier)
Power
Supply
Primary
Saturating
Winding
Secondary
LOAD

Figure 16. Magnetic amplifier with ultracarrier frequency "agitating" means for hysteresis-effect reduction

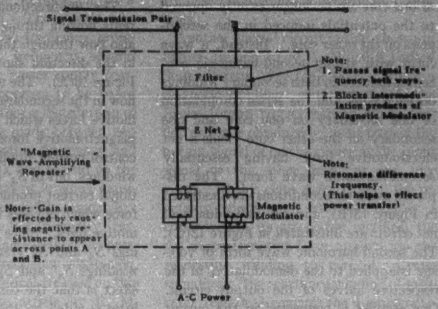

Signal Transmission Pair
A
B
Filter
Note:
1. Passes signal frequency both ways.
2. Blocks intermodulation products of Magnetic Modulator
Z Net
"Magnetic
Wave-Amplifying
Repeater"
Note:
Resonates difference frequency.
(This helps to effect power transfer)
Note: Gain is effected by causing negative resistance to appear across points A and B.
Magnetic
Modulator
A-C Power

Figure 17. Series negative-resistance type magnetic amplifier

If the control signal is zero, equal and opposite potentials of the same frequency as the power-supply frequency are induced in the secondary windings. Equal and opposite currents thus are caused to flow through the rectifiers and through the two halves of the output device; the net output signal is consequently zero.

Assume that a positive control current is made to flow in the saturating winding. The wave shapes of the voltages induced in the secondary windings will now assume the forms shown in Figure 12(C) because of the displaced excitation conditions illustrated in Figure 12(B). The rectifier associated with secondary S_A, will pass positive half-cycles of current, and the rectifier associated with secondary S_B will pass negative half-cycles. Because these rectifiers exhibit a considerably reduced impedance when the applied forward voltage is increased, a notably greater current flows through the rectifier which is conducting than through the rectifier which is not conducting. Due both to the distortion in secondary voltage wave form and to the nonlinear resistance characteristic of the rectifiers, therefore, a considerable unbalance of voltage drop is made to occur between the two halves of the output device.

The unidirectional currents flowing through the output devices also flow in secondary windings S_A and S_B in unbalanced magnitudes such that a net unidirectional magnetomotive force exists in a direction aiding the original control magnetomotive force. This amounts to positive internal feedback which increases the amplifier output.

Negative control-current signals have similar, though opposite, effects.

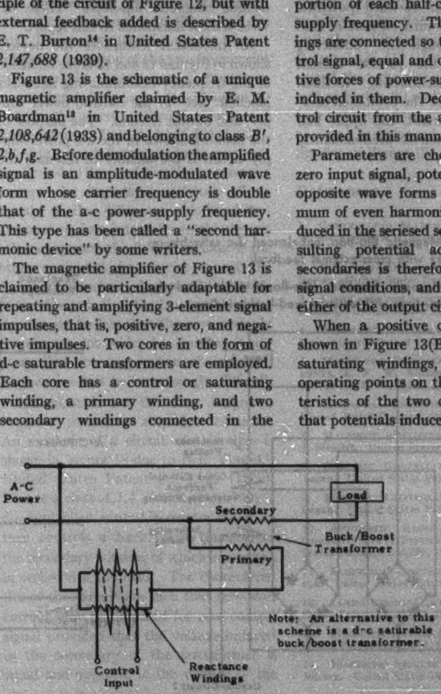
A magnetic flip-flop based on the principle of the circuit of Figure 12, but with external feedback added is described by E. T. Burton[14] in United States Patent *2,147,688* (1939).

Figure 13 is the schematic of a unique magnetic amplifier claimed by E. M. Boardman[15] in United States Patent *2,108,642* (1938) and belonging to class B', *2,b,f,g*. Before demodulation the amplified signal is an amplitude-modulated wave form whose carrier frequency is double that of the a-c power-supply frequency. This type has been called a "second harmonic device" by some writers.

The magnetic amplifier of Figure 13 is claimed to be particularly adaptable for repeating and amplifying 3-element signal impulses, that is, positive, zero, and negative impulses. Two cores in the form of d-c saturable transformers are employed. Each core has a control or saturating winding, a primary winding, and two secondary windings connected in the equivalent of a push-pull system. A-c power is introduced via the two primary windings, and the power-supply frequency is considerably higher than any component of the input signal frequency. An essential feature of the primary excitation of this circuit is that it must be sufficient to saturate the cores partially during a portion of each half-cycle of the power-supply frequency. The saturating windings are connected so that, with zero control signal, equal and opposing electromotive forces of power-supply frequency are induced in them. Decoupling of the control circuit from the a-c power supply is provided in this manner.

Parameters are chosen so that, with zero input signal, potentials of equal and opposite wave forms containing a minimum of even harmonic distortion are induced in the seriesed secondaries. The resulting potential across the seriesed secondaries is therefore zero under zero signal conditions, and no current flows in either of the output circuits.

When a positive current impulse, as shown in Figure 13(B), is applied to the saturating windings, however, the net operating points on the magnetic characteristics of the two cores are shifted so that potentials induced in the secondaries


A-C
Power
Load
Secondary
Buck/Boost
Transformer
Primary
Control
Input
Reactance
Windings

Note: An alternative to this scheme is a d-c saturable buck/boost transformer.

Figure 18. Buck/boost transformer controlled by d-c saturable reactor

of one core are no longer exactly opposed to the potentials induced in the secondaries on the other core. Instead, a wave form with strong second-harmonic content is induced in both secondary windings on each core, and the series combination of one secondary on one core and one secondary on the other core yields a net electromotive force having essentially second-harmonic wave form. The displaced excitation conditions are illustrated in Figure 13(B), and the individual and net effects are illustrated in Figure 13(C). The second-harmonic wave form of voltage is applied to the demodulators in the respective halves of the output circuit. One rectifier is connected in the output circuit comprising windings S_A and S_B in such a manner that it will conduct under the influence of high-potential positive peaks induced in this circuit. Another rectifier is oppositely connected in the output circuit comprising windings S_A' and S_B' in a direction so that it will conduct under the influence of the negative peaks, but not the positive peaks. The rectifier elements have nonlinear characteristics (copper-oxide rectifiers, for example) in which the forward impedance decreases rapidly as forward potential increases. The forward current is thus much greater through the conducting rectifier than is the reverse current through the nonconducting rectifier. Therefore, because of the combined characteristics of the circuit, the upper half of the output device is predominantly energized.

The unidirectional demodulated currents flowing through the output devices also flow through the secondary windings in the direction shown by the arrows in Figure 13(A). The unidirectional current flow in the secondaries produces magnetomotive forces which in S_A and S_B aid the magnetomotive force due to the positive control current applied to the saturating windings, but in S_A' and S_B' the unidirectional current produces a magnetomotive force which opposes the control magnetomotive force. Since the current in windings S_A and S_B greatly exceeds that in windings S_A' and S_B', however, the net effect is that the control magnetomotive force is aided by the feedback magnetomotive force which results from the flow of demodulated output current in the secondary windings. This action amounts to positive internal feedback and operates to increase the output of the amplifier.

Negative current impulses applied to the control circuit have similar though opposite effects.

A group of devices similar to that shown in Figure 13 are the subjects of E. T. Burton's United States Patents *2,147,688* (1939)[14] and *2,164,383* (1939)[15] which include magnetic flip-flops. These devices are not illustrated in this report.

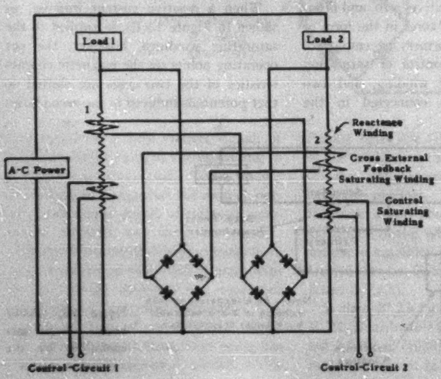
An important feature of the devices shown in Figures 12 and 13 is the inclusion of the frequency-discriminating networks labelled "net" in the control, power supply, and output circuits. The purpose of the network shown in the a-c power-supply circuit is presumably to resonate the primary circuit to the frequency of the a-c power supply (a series capacitor for the same purpose is included in the reactor circuit of Figure 6). The purpose of the network in the control circuit is to prevent the power-supply frequency harmonic voltages induced in the saturating windings from affecting the control circuit, and to prevent changes in the impedance of the control line from affecting the operation of the amplifier. The network in the output circuit, Figure 13, is to improve the wave forms of the amplified signals which are transmitted to the output device.

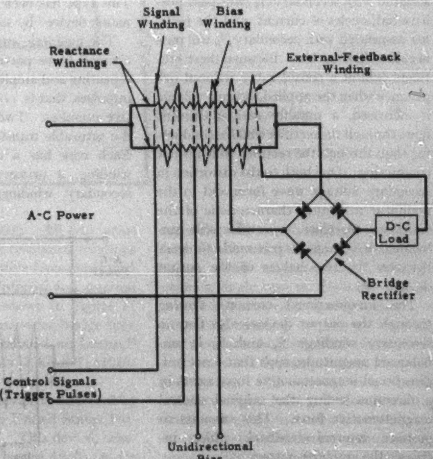
A class *A,1,g* magnetic amplifier, which is possibly unique in the art, is shown in Figure 14. Described by H. D. Middel[16] in United States Patent *2,388,070* (1945), this device comprises a d-c saturable reactor in which the saturating flux produced by the control winding causes a second harmonic electromotive force of power-supply frequency to be induced in the load windings. Under the influence of control saturation, an asymmetrical voltage is caused to act across a nonlinear resistor seriesed with the load winding of each stage. This departure from symmetry causes the nonlinear resistor to act as a rectifier which introduces a d-c component in the load circuit. A minute change of d-c signal applied to the saturating winding of the reactor will thus cause a much larger change in d-c load current.

The symmetrical nonlinear impedance elements in this device have no unidirec-

Figure 19 (below). Magnetic flip-flop, unbalanced d-c saturable-reactor type with cross-external feedback

Figure 20 (right). Magnetic amplifier (flip-flop), single-sided d-c saturable reactor type with external feed-back winding





tional property or rectifying tendency per se. For example, they may be either Thyrite or parallel back-to-back rectifiers such as copper-oxide rectifiers. The basic elements of the amplifier are thus a symmetrical current-distorting impedance, an asymmetrical current-distorting impedance with polarizing means, an a-c power supply, and a d-c load circuit. It may be that this device should be considered as a general case within which the conventional internal-feedback magnetic amplifier, as shown in Figure 10, is a special case. The reader is referred to the patent for a detailed description of this device.

Important off-resonant and ferroresonant means for increasing the gain of magnetic amplifiers have been described from an early date. Figure 15 shows an example of the use of such sensitivity increasing means in a (class *A,1,a,d,e*) voltage-regulating device due to L.R. Runaldue,[17] United States Patent *2,278,151* (1942). A parallel inductance-capacitance off-resonant circuit is used for load-current control. The range of impedance obtainable across the terminals of the parallel inductance-capacitance circuit as it approaches resonance is considerably greater than the range which might be obtained by use of the saturable reactor alone. The impedance of this network is changed by varying the inductance of the d-c saturable reactor. If the inductive reactance of the reactor approaches the capacitive reactance of the paralleled capacitor, the network will approach a resonant condition and the impedance will approach a very high value. If the inductive reactance of the reactor is either increased or decreased from the resonant value, the impedance of the network will decrease.

The sensitivity of the control circuit in Figure 15 may be increased by the inclusion of a ferroresonant combination-comprising a series-connected capacitor, resistor, and self-saturating inductance. By use of the ferroresonant combination, the current in the control circuit under the influence of a varying voltage excitation may be made to increase at a much greater rate than the excitation voltage. In other words, the ferroresonant combination amplifies the current or voltage variations occurring in the circuit from which it is excited.

Figure 16 illustrates an unusual and little-exploited means for increasing the gain of a magnetic amplifier by altering the hysteresis effects of the core material. In this class *B′,1,c,h*, device, due to R.V. L. Hartley,[18] United States Patent *1,287-982* (1918), the inventor describes means of altering the hysteresis effects in the core material by the application of a special agitating field. The core is subjected to a weak high-frequency cross-magnetization at a frequency several times that of the a-c power supply (carrier). This patent appears to be based first on the principle that a distribution of hysteresis loss can be made between two frequencies (the agitating frequency and the carrier frequency), and second on the corollary principle that at low flux densities the dynamic incremental permeability of the material at the low frequency (carrier) is considerably increased by the presence of the higher frequency excitation. For a pertinent theoretical treatise on this phenomenon, see the article by R. M. Kalb and W. R. Bennett.[19] See also G.W. Elmen and C. F. Ort,[20] United States Patent *1,544,381*, 30 June, 1925, which applies the phenomenon.

A significant and novel type of magnetic amplifier is the negative resistance type shown in Figure 17 and designated as class *C,1,f*. This device was described by E. Peterson[21] in United States Patent *1,884,844* (1932). The operation of this circuit depends upon the intermodulation of power-supply and signal frequencies in such manner that energy is transferred from the power source to the signal circuit. The energy transfer results when a difference side band of the two frequencies (for example, twice power frequency minus signal frequency) is permitted to flow in the circuit. See K. Heegner,[22] United States Patent *1,656,195* (1928). Other intermodulation products must not be allowed to flow and thereby to consume power. An impedance network allows signals of the difference side-band frequencies to flow internally in the device. Filters suppress the flow of signals of the other generated frequencies except the signal frequencies. A general limitation of the device is that the bandwidth of the signal must be less than the power-supply frequency; for wider bandwidths it appears that the suppression of unwanted frequencies will be increasingly difficult.

Classification 4 of the chart designates "special magnetic amplifier circuits." An example of a circuit of this class is shown in Figure 18 due to H.E. Young,[23] United States Patent *2,154,020* (1939). This is a class *A,1,4*, circuit comprising a d-c controlled saturable reactor which in turn controls a buck/boost transformer the secondary winding of which is inserted in series with the load. The characteristics of this circuit are such that zero load current may be obtained for zero control signal provided that the voltage induced in the secondary of the transformer is equal and opposite to the voltage of the a-c power source under zero-signal conditions.

Another special circuit is that shown in Figure 19 due to M. L. Wood,[24] United States Patent *2,524,154* (1950). This is a magnetic-amplifier counterpart of the Eccles-Jordan trigger pair in which flip-flop action is obtained by using part of the output signal of one amplifier as the input signal to the other amplifier and vice versa. In this class *A,4,a* circuit, the reactance winding of saturable reactor *1* controls (via its associated rectifier) the d-c magnetomotive force applied via the saturating winding to core *2*. Correspondingly, the reactance winding of saturable reactor *2* controls the d-c magnetomotive force applied via the saturating winding to core number *1*. This flip-flop may be said to be a "cross external-feedback type" of magnetic amplifier. The state of this flip-flop may be changed by applying unidirectional magnetomotive forces by means such as the control windings shown.

Another type of magnetic flip-flop is shown in Figure 20. This is a single-sided magnetic amplifier similar to that shown in Figure 5, but with a sufficient amount of external feedback to cause instability. This class *A,1,a,c* device is described by FitzGerald[25] in United States Patent *2,027,312* (1936). It has been suggested that the signal-sided flip-flop circuit of Figure 20 may be used to advantage in that it does not require equal loads as does the balanced Eccles-Jordan type flip-flop of Figure 19.

References

1. CONTROLLING ALTERNATING CURRENTS, **E. F. W. Alexanderson.** United States Patent *1,206,643*, 1916.
2. ELECTRON DISCHARGE APPARATUS, **A. E. Bower.** United States Patent *2,230,558*, 1941.
3. REGULATOR FOR ALTERNATING CURRENTS, **E. R. Stockle.** United States Patent *1,376,978*, 1921.
4. RAILWAY SIGNALLING APPARATUS, **C. M. Hines.** United States Patent *2,215,820*, 1940.
5. MEANS FOR CONTROLLING ALTERNATING CURRENTS, **E. F. W. Alexanderson.** United States Patent *1,328,797*, 1920.
6. MAGNETIC AMPLIFIER, **A. S. FitzGerald.** United States Patent *2,464,639*, 1949.
7. MAGNETIC AMPLIFIER AND SELF-MODULATOR, **J. Slepian.** United States Patent *1,645,302*, 1927.
8. ELECTRIC CONTROLLING APPARATUS, **F. G. Logan.** United States Patent *1,997,179*, 1935.
9. ELECTRIC CONTROLLING APPARATUS, **F. G. Logan.** United States Patent *2,126,790*, 1938.
10. AUTOMATIC REGULATING APPARATUS FOR CURRENT SUPPLY SYSTEMS, **S. M. Hanley.** United States Patent *2,144,290*, 1939.
11. ELECTRICAL CONTROL SYSTEM, **L. W. Thompson.** United States Patent *1,844,704*, 1932.
12. LOAD CURRENT CONTROL, **Uno Lamm.** United States Patent *2,403,891*, 1946.
13. MAGNETIC DEVICE, **E. M. Boardman.** United States Patent *2,108,642*, 1938.
14. MAGNETIC IMPULSE STORING DEVICE, **E. T. Burton.** United States Patent *2,147,688*, 1939.

15. Magnetic Device, E. T. Burton. United States Patent *2,164,383*, 1939.

16. Electromagnetic Apparatus, H. D. Middel. United States Patent *2,388,070*, 1945.

17. Regulating Apparatus, L. R. Runaldue. United States Patent *2,278,151*, 1942.

18. Modulating System, R. V. L. Hartley. United States Patent *1,287,982*, 1918.

19. R. M. Kalb, W. R. Bennett. *Bell System Technical Journal* (New York, N. Y.), volume XIV, April 1935, pages 322–59.

20. Method and System for Amplifying Variable Currents, G. W. Elmen, C. F. Ort. United States Patent *1,544,381*, June 30, 1925.

21. Magnetic Wave-Amplifying Repeater, E. Petterson. United States Patent *1,884,844*, 1932.

22. Arrangement for the Generation of Oscillations, K. Heegner. United States Patent *1,656,195*, 1928.

23. Constant Voltage System, H. E. Young. United States Patent *2,154,020*, 1939.

24. Electrical Trigger, M. L. Wood. United States Patent *2,524,154*, 1950.

25. Magnetic Amplifying and Control System, A. S. FitzGerald. United States Patent *2,027,312* 1936.

No Discussion

The Figure of Merit of Magnetic Amplifiers

J. T. CARLETON
ASSOCIATE MEMBER AIEE

W. F. HORTON
ASSOCIATE MEMBER AIEE

REGULATING systems are a combination of devices whereby a large amount of energy is controlled by a small amount of energy. Large amounts of heat can be controlled by a small amount of thermal energy in a thermocouple. A feeble voltage can regulate the speed of a tandem mill; a tiny magnet governs the voltage of a large alternator; and a microwatt radar signal can guide a missile at supersonic speeds. It follows that power amplifiers play a major role in regulating systems; therefore, some of the principle characteristics of proportional power amplifiers will be examined.

Characteristics of Power Amplifiers

POWER AMPLIFICATION

Of primary importance in power amplifiers is the ratio of power output to power input of the amplifier, namely the power amplification. It is fundamental that the system power amplification must be such that the output power can be controlled by the available controlling energy; hence, the system power amplification is determined by the power output required, the power level of the detecting device, the nature of the disturbances to which the system is subjected, and the degree of regulation required.

RESPONSE

The rapidity with which an amplifier responds to an input signal is also of importance in determining the performance of regulating systems. Associated with the output power of a power amplifier is the storage of a certain amount of energy in the control elements of the amplifier. The response of the thermocouple is determined by the amount of heat stored in bringing it up to the proper temperature. Amplifiers utilizing magnetic fields store energy in the form of flux. A Silverstat contains energy stored in a stressed spring and moving inertia; and vacuum tubes are plagued with electrostatic energy stored in the interelectrode capacities. The phenomenon of changing the amount of stored energy for a change of input power prevents an instantaneous change in power output to conform to a change in input. This delay in response is frequently expressed as the time constant of the amplifier.

GAIN

A third important characteristic of power amplifier is gain, or, in the case of electric power amplifiers, voltage amplification. System gain governs the regulator's stiffness and, in conjunction with the amplifier's delays, determines the system response and stability. Unlike vacuum tubes, whose voltage gains are fairly well fixed, magnetic amplifiers, both static and rotating, enjoy flexibility, the voltage gains obtainable being limited mainly by the maximum and minimum wire sizes that may be used in the control and output circuits.

In designing regulating systems it is necessary to be able to evaluate the power amplification, delay, and gain of the amplifiers involved. In static and rotating magnetic amplifiers, because the gain is flexible the power amplification and delays define the capability of the amplifier. For this purpose the "figure of merit" of the amplifier has been defined as the ratio of power amplification to time constant. In this paper, it is demonstrated that a figure of merit of static magnetic amplifiers so defined exists, that it corresponds to one-half the ratio of power output to stored magnetic energy, and that it is calculable from the configuration of the amplifier. A method of designing magnetic amplifiers with high figure of merit also is presented. Methods of calculating the figure of merit of d-c rotating power amplifiers is presented elsewhere.

The Equivalent Circuit and Transfer Function of Magnetic Amplifiers

The performance of magnetic amplifiers in regulating systems may be expressed in the form of a transfer function. This function may be obtained by linearizing the magnetic amplifier and making certain simplifying assumptions. The simplified device may be understood and analyzed more easily by examining its equivalent circuit.

Therefore, a transfer function and an equivalent circuit for a self-saturating magnetic amplifier will be derived. The operation of the self-saturating magnetic amplifier has been described in the literature.[1] Briefly a control winding current establishes flux conditions in the magnetic structure, the flux conditions at the start of any positive half-cycle or supply voltage determining the load voltage. The control ampere-turns may take a series of valves, as shown in Figure 1, over a useful range *OB*, which determines the range of control ampere-turns necessary for utilization of the amplifier over its full range of output. This range of ampere-turns can be related to core properties, as illustrated in Figure 1, by the superimposed amplifier control characteristic and the operating curve of flux density versus field intensity.[2] Effects of rectifier leakage are neglected.

The relationship between control ampere-turns and the magnetic material and structure is

$$H_i = \frac{0.4\pi N_i I_i}{l} \text{ (oersteds)} \tag{1}$$

when l is the effective mean length of the core magnetic path in centimeters. The change in control H corresponding to a change in control ampere-turns OB is

$$\Delta H_i = \frac{0.4\pi N_i \Delta I_i}{l} \text{ (oersteds)} \tag{2}$$

ΔH_i is dependent on core material, lamination thickness, and supply fre-

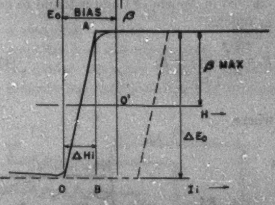
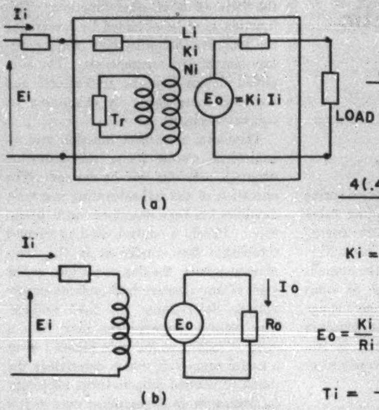


Figure 1. Operating β-H curve with superimposed control characteristic

—— Idealized control characteristic and left flank and top of operating β-H curve
—— Actual control characteristic
– – Right flank and bottom of operating β-H curve

Paper 52-234, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted March 25, 1952; made available for printing May 14, 1952.

J. T. CARLETON and W. F. HORTON are both with Westinghouse Electric Corporation, East Pittsburgh, Pa.



$$Li = \frac{Ni\,\Delta\phi}{\Delta Ii} \times 10^{-8} = \frac{4\pi Ni^2 A_c(2\beta m)}{\ell\,\Delta Hi} \times 10^{-8}$$

$$Ki = \frac{\Delta Eo}{\Delta Ii} = \frac{4(.4\pi)\,fNoNiAc\beta m}{\ell\,\Delta Hi} \times 10^{-8}$$

$$Ki = 2f\left[\frac{No}{Ni}\right] Li$$

$$Eo = \frac{Ki}{Ri}\left[\frac{1}{1+pTi}\right] Ei$$

$$Ti = \frac{Li}{Ri}$$

Figure 2(A). Equivalent circuit of self-saturating magnetic amplifier linearized over the curve OA. (B). Simplified equivalent circuit

quency. It is the change in control oersteds required to change the flux at the beginning of a positive half-cycle of the supply voltage over the left flank of the operating hysteresis loop.

The relationship between ΔE_o and β_m is

$$\Delta E_o = 4N_o f A_c \beta_m \times 10^{-8} \qquad (3)$$

where ΔE_o is the average value of the a-c supply voltage and the maximum output voltage of a single-phase amplifier.

Figure 2(A) is the equivalent circuit of the self-saturating magnetic amplifier shown in Figure 6. Figure 2(B) is a simplified equivalent circuit derived by lumping the resistances of the control circuit and the output circuit respectively and by neglecting the inductively coupled delay produced by the eddy currents and bias windings T_r. The transfer function is derived in Appendix I and is found to be as follows

$$E_o = \frac{K_i}{R_i}\,\frac{1}{1+p\dfrac{L_i}{R_i}}\,E_i \qquad (4)$$

where

E_o = the internal output voltage
E_i = input voltage
$K_i = \dfrac{\partial E_i}{\partial I_i}$
L_i = control current
R_i = control circuit resistance
L_i = control circuit inductance

Figure 3 is the equivalent circuit of a simple series-reactor-type magnetic amplifier over linear range of operation. The transfer function is derived in Appendix II and is found to be as follows

$$E_o = \frac{1}{1+p\dfrac{L_i}{R_i}}\left(\frac{K_i}{R_i}E_i - \frac{N_o}{N_i}K_i I_o\right) \qquad (5)$$

If the amplifier is loaded into a resistive load R_o and $(N_o/N_i)K_i \gg R_o$ which is generally the case when high performance magnetic-material is used, the following transfer function applies

$$E_o = \frac{N_i}{N_o}\,\frac{R_o}{R_i}\,\frac{1}{1+p\dfrac{N_i}{N_o}\,\dfrac{R_oL_i}{R_iK_i}}\,E_i \qquad (6)$$

When the amplifier is loaded into an inductive load (see Figure 4) the advantage of the negative current feedback is utilized, for the transfer function then becomes

$$E_o = \frac{R_o}{R_i}\,\frac{N_i}{N_o} \times \frac{1+p\dfrac{L}{R_o}}{1+p\,N_iR_o\left(\dfrac{L_i}{R_i}+\dfrac{L}{R_o}\right)+p^2\,\dfrac{N_iR_o}{N_oK_i}\,\dfrac{L_iL}{R_iR_o}}\,E_i \qquad (7)$$

The equivalent circuit of the parallel connected magnetic amplifier is shown in Figure 5 in which the delay produced by the a-c windings is included.

Energy Storage and the Figure of Merit

The figure of merit of a magnetic amplifier is defined as the ratio of power amplification to time constant

$$m = A_p/T \text{ per second} \qquad (8)$$

or alternately

$$m = A_p/T \text{ per cycle} \qquad (9)$$

where A_p is the amplifier power amplification and T is the amplifier time constant expressed in seconds, equation 8, or in cycles of the supply frequency, equation 9. Power amplification can be expressed

$$A_p = \frac{\Delta P_o}{\Delta P_i} = \frac{\Delta I_o^2 R_o}{\Delta I_i^2 R_i} \qquad (10)$$

the ratio of a change in output power to a change in input or control power. (For notation see Appendix III, Figure 6, and Figure 7.) The time constant of a magnetic amplifier is determined by the effective inductance of its control winding and the resistance of its control circuit which gives

$$T = L_i/R_i \text{ (seconds)} \qquad (11)$$

From equations 8, 10, and 11 is seen that

$$m = \frac{\Delta P_o}{L_i \Delta I_i^2} \text{ per second} \qquad (12)$$

or that the figure of merit of a magnetic amplifier is determined by the ratio of the change in amplifier power output to twice the change in energy stored in the amplifier control circuit. Equation 12 presents a clear physical picture of the operation of the power amplifier—a change in

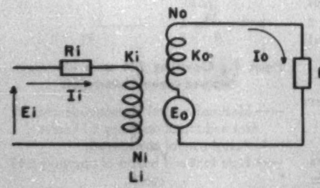
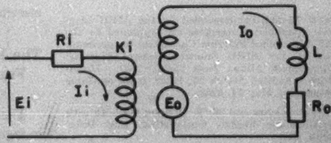


Figure 3 (left). Equivalent circuit of series-connected reactor magnetic amplifier with resistive load

Figure 4 (right). Equivalen circuit of series-connected reactor magnetic amplifier with inductive load



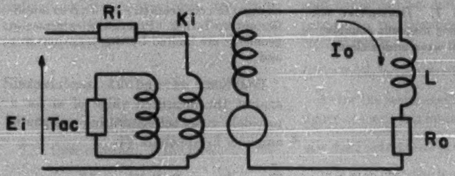


Figure 5 (left). Equivalent circuit of parallel-connected reactor magnetic amplifier with inductive load

control energy storage must accompany any change in output power.

Self-Saturating Magnetic Amplifier

Self-saturation in magnetic amplifiers refers to the circuitry of a group of basic single-phase and multiphase amplifiers. Amplifiers of this type have been described at some length in the literature. It has been customary in the literature to treat self-saturating amplifier circuits as simple reactor circuits with the addition of almost 100-per-cent positive feedback. When this positive feedback is added to a simple reactor circuit, the cyclic demagnetization of the reactors by the supply voltage no longer occurs and the control characteristic of the amplifier is primarily dependent on the characteristics of the reactor core material. The significant advance of the self-saturating circuit over the simple reactor circuit is in the ratio of power gain to amplifier time constant. A simple reactor circuit with a value of 3 for this ratio may have a ratio of 500 with the addition of self-saturating rectifiers. Since the self-saturating circuits do depend primarily on the characteristics of the reactor core material, they may be considered the basic magnetic amplifier circuits. In the same sense the simple reactor circuits can be considered as self-saturating circuits employing almost 100-per-cent negative current feedback.

The single-phase self-saturating circuits, see Figure 6, (the doubler, bridge circuit, and center-tapped transformer circuit) can be grouped and said to have very similar steady-state characteristics.[3] Their output power capabilities and required control ampere-turns are the same. The transient characteristics of this group differ in residual time constant which is dependent on specific circuitry. This is a time constant not associated with the amplifier control circuit and as such is not amenable to control circuit "forcing."

The 3-phase self-saturating circuits, see Figure 7, are combinations of single-phase circuits. The steady-state and transient characteristics of these circuits are dependent on the characteristics of their single-phase components and, as such, generalizations which apply to the single-phase component circuits will apply to the 3-phase circuits.

The figure of merit of a magnetic amplifier circuit

$$m = A_p/T \text{ per cycle} \tag{9}$$

is a constant for a given amplifier operating into a fixed load when residual time constant can be neglected. This expresses the experimentally proved fact that amplifier power gain and time constant are inversely related. One error in this relationship lies in the amplifier residual time constant. The residual time constant of the single-phase bridge and center-tapped transformer circuits is of the order of 0.5 cycle[4] of the supply frequency and of the order of 1 cycle for the doubler circuit. A second source of error lies in the phenomenon of commutation[5] at high values of control circuit resistance per control winding turns squared. When the amplifier operating time constant is high with respect to residual time constant, equation 9 defines a figure of merit which is substantially constant over a wide range of time constants and power amplifications.

To make the figure of merit concept useful, amplifier design conditions will be derived from it.

Single-Phase Magnetic Amplifier

PROPORTIONING OF REACTOR WINDOW AREA

The figure of merit of a magnetic amplifier bears a direct relationship to the relative proportions of reactor window area which can be devoted to the load winding and the window area which must be devoted to the control winding for a given measure of control. This relationship can be derived from the definition of m.

The control circuit time constant of a single-phase self-saturating magnetic amplifier is one-half the average voltage gain on a common turns base.[6]

$$T = 1/2A_e \text{ (cycles)} \tag{13}$$

From equations 9 and 13

$$m = 2A_I \text{ per cycle} \tag{14}$$

where A_I is amplifier average current gain on a common turns base and is measurable directly from the amplifier control characteristic.

From equation 14

$$\frac{\Delta N_o I_o}{\Delta N_i I_i} = \frac{m}{2} \tag{15}$$

The ratio of average output current to rms single reactor current is $1/\sqrt{2\mu}$ at maximum amplifier output. μ is dependent on load inductance and varies from 0.95 to 1 as the load varies from resistive to highly reactive. Assuming that output and control windings are worked at equal current densities, $m/2\sqrt{2\mu}$ is the ratio of load copper per core to control copper per core and is thus roughly equivalent to the ratio of window areas required by load to control winding. To obtain orders of magnitude, measured values of m are of the order of 200 to 1,500 in amplifiers employing square loop core materials, depending on actual core dimensions. For amplifiers of this type the reactor window area required by the control winding for full control is almost negligible as compared to that required by the load winding.

RELATION OF CORE DIMENSIONS AND CHARACTERISTICS TO FIGURE OF MERIT

Amplifier core dimensions and core material characteristics bear a very important relationship to the figure of merit. It is desirable in many cases to design the amplifier core and windings, the only independent valuables that the amplifier designer has readily at hand, to conform

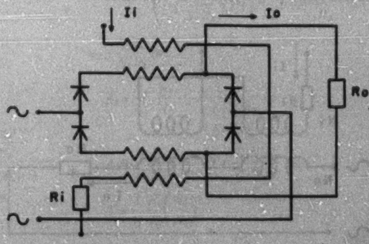


Figure 6 (right). Single-phase self-saturating magnetic amplifier—bridge circuit

R_o = lumped load circuit resistance

R_i = lumped control circuit resistance

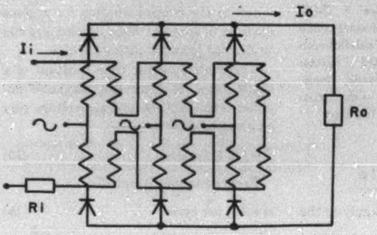


Figure 7. Three-phase self-saturating magnetic amplifier—full-wave bridge circuit

with this relationship—in other words, to maximize the amplifier figure of merit.

Corresponding to ΔH_t change in control oersteds, load current changes from approximately zero, magnetizing current only, to the value corresponding to full amplifier saturation.

$$\Delta I_o = \frac{\Delta E_o}{R_o} - I_o \text{ mag} \qquad (16)$$

or

$$\Delta I_o \approx I_o \text{ max} \qquad (16A)$$

where ΔE_o is the average value of the supply voltage and R_o is the total output circuit resistance.

From equations 2 and 14

$$m = 2 I_o \text{ max } N_o \frac{0.4\pi}{\Delta H_t l} \text{ per cycle} \qquad (17)$$

$I_o N_o$ can be expressed in terms of core dimensions from the considerations outlined in the previous section on "Proportioning of Reactor Window Area." In Appendix IV an expression is derived from m in terms of readily determinable core constants and dimensions. This expression is

$$m = \frac{2\sqrt{2\mu} s A_w}{\rho} \frac{0.4\pi}{\Delta H_t l} \text{ per cycle} \qquad (18)$$

As defined in Appendix I,

s is the load winding space factor
A_w is the window area of the core in circular mils
ρ is the circular mils per rms amperes for the load winding at full output

When applied to a particular amplifier circuit employing a particular core structure, equation 18 shows the required design considerations. Equation 18 is applied to two important amplifier core structures, the rectangular and toroidal structures in Appendix V. Some immediate generalizations are possible:

1. m is independent of the saturation flux density of the core material.
2. m is independent of stack height for stacked cores or strip width for toroidal cores.
3. m is maximized when core build up is minimized.
4. m increases as a linear dimension of the core window—in toroidal cores as the inside diameter is increased, in rectangular cores as window width and length is increased.
5. Since a circle is capable of a greater ratio of area to circumference than a rectangle, toroidal construction will give the maximum value of m, other factors being equal. The comparative ratio is $4/\pi = 1.3/1$, comparing a toroidal with a square core with equal control requirements for the theoretical case of zero core build-up.
6. m is maximized for maximum load winding space factor and maximum working current density of load copper.
7. m is inversely proportional to ΔH_t.

These generalizations of course are dependent on the assumptions of the derivation of equation 18. These assumptions are:

1. Minimum amplifier output is essentially zero when compared with maximum output.
2. Amplifier control characteristic is linear from minimum output to maximum output.
3. ρ and s do not change appreciably for changes in core geometry for a given amplifier output rating. Practically extremes in geometry are limited by the variations in ρ and s.

Equations 14 and 18 are checked against the measured value of m for a toroidal core self-saturating magnetic amplifier in Appendix VI.

Three-Phase Full-Wave Magnetic Amplifier

PROPORTIONING OF REACTOR WINDOW AREA

The control circuit response time of a 3-phase full-wave magnetic amplifier is

$$T = A_s/3 \text{ (cycles)} \qquad (19)$$

This expression can be used with equation 2 to give

$$m = 3A_I \text{ per cycle} \qquad (20)$$

where A_I is again amplifier average current gain on a common turns base, or

$$\frac{\Delta N_o I_o}{\Delta N_t I_t} = \frac{m}{3} \text{ per cycle} \qquad (21)$$

The ratio of average output current to rms single reactor current for the 3-phase bridge circuit is $1/\sqrt{3}$. Assuming again that output and control windings are worked at equal current densities, $m/3\sqrt{3}$ is the required ratio of load copper per core to control copper per core.

Similarly, ratios of output winding copper area to control winding copper area can be calculated for other 3-phase circuits.

RELATION OF CORE DIMENSIONS AND CHARACTERISTICS TO FIGURE OF MERIT

An expression for m similar to that of equation 18 can be derived for the full-wave 3-phase amplifier circuit. This expression is

$$m = \frac{3\sqrt{3} s A_w}{\rho} \frac{0.4\pi}{\Delta H_t l} \text{ per cycle} \qquad (22)$$

where the terms are as defined before. Similarly an expression for m can be derived for other 3-phase circuits. The gen-

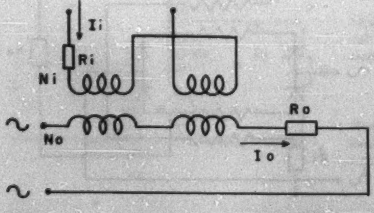
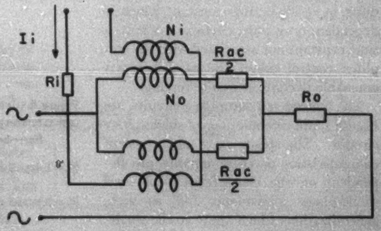


Figure 8 (left). Series-connected reactor magnetic amplifier

Figure 9 (right). Parallel-connected reactor magnetic amplifier



eralizations drawn previously are again true within the limits of the assumptions of the derivation.

Simple Reactor Magnetic Amplifier

SERIES-CONNECTED REACTORS

The time constant of the magnetic amplifier consisting of two series-connected reactors controlling a resistive load, as shown in Figure 8, has been derived by Storm.[7] The time constant in cycles is given by

$$T=\frac{1R_o}{4R_i}\frac{N_i^2}{N_o^2}\text{ per cycle} \tag{23}$$

neglecting residual delay associated with the load circuit and eddy currents. Power gain of the amplifier can be derived from the equal ampere-turns relationship which governs simple reactor control characteristics over the control range. This relationship can be expressed as

$$N_oI_o=N_iI_i\text{ (ampere-turns)} \tag{24}$$

$$\frac{(I_o)^2}{(I_i)^2}=\frac{(N_i)^2}{(N_o)} \tag{25}$$

and

$$A_p=\frac{(I_o)^2R_o}{(I_i)^2R_i}=\left(\frac{N_i}{N_o}\right)^2\frac{R_o}{R_i} \tag{26}$$

From equation 23 the figure of merit can be obtained

$$m=\frac{A_p}{T}=\frac{\frac{R_o}{R_i}\left(\frac{N_i}{N_o}\right)^2}{1/4\frac{R_o}{R_i}\left(\frac{N_i}{N_o}\right)^2}=4\text{ per cycle} \tag{27}$$

Thus the figure of merit *m* for a simple series reactor magnetic amplifier, when operating into a resistive load, is a constant numerically equal to 4 per cycle. When operating into an inductive load the negative current feedback of the reactor should be considered as outlined in the consideration of equivalent circuits.

PARALLEL-CONNECTED REACTORS

The time constant of the magnetic amplifier consisting of two parallel-connected reactors controlling a resistive load, as shown in Figure 9, is not easily expressible in terms of load and control circuit resistances and turns as in the case of the series-connected reactors. The time constant in cycles can be expressed[8] as

$$T=\frac{R_o}{R_i}\left(\frac{N_i}{N_o}\right)^2(1+\alpha)\text{ per cycle} \tag{28}$$

where α is the ratio R_{ac}/R_o, again neglecting residual delay.

R_{ac} is defined as the sum of the resistances of reactor load windings. Power gain again can be derived from the ampere-turns relationship over the useful control region.

$$I_iN_i=1/2I_oN_o\text{ (ampere-turns)} \tag{29}$$

As for the series-connected reactors, the figure of merit can be calculated as

$$m=A_p/T=4/(1+\alpha)\text{ per cycle} \tag{30}$$

It is seen that the figure of merit for the parallel connected reactor amplifier is always less than for the series-connected reactor amplifier by the ratio $1/(1+\alpha)$.

Conclusions

FIGURE OF MERIT SELF-SATURATING MAGNETIC AMPLIFIER

1. The figure of merit and equivalent circuits are determinable from core geometry and characteristics.
2. The figure of merit is determinable experimentally by static test.
3. Correlation of figure of merit as determined by the methods described in "Proportioning of Reactor Window Area" and "Relation of Core Dimensions and Characteristics to Figure of Merit" check very closely.

FIGURE OF MERIT SIMPLE REACTOR MAGNETIC AMPLIFIER

1. The figure of merit for resistive load is independent of core geometry and within wide limits independent of core characteristics.
2. Core geometry and characteristics determine current forcing effect when operating into inductive loads.
3. The simple reactor amplifier is useful when output current independent of output voltage is desired.

Appendix I. Transfer Function of Self-Saturating Magnetic Amplifier

Refer to Figure 2

$$E_i=I_iR_i+p\frac{L_i}{K_i}E_o$$

$$E_o=K_iI_i$$

$$E_o=\frac{K_i}{R_i}\frac{1}{1+p\frac{L_i}{R_i}}E_i$$

where

$$K_i=\frac{\partial E_o}{\partial I_o}=\frac{\Delta E_o}{\Delta I_o}=\frac{4N_ofA_c\beta_m\times10^{-8}}{l\Delta H_i/0.4\pi N_i}$$

$$=\frac{1.6\pi fN_oN_iA_c\beta_m\times10^{-8}}{l\Delta H_i}$$

$$L_i=\frac{N_i\phi}{I_i}\times10^{-8}=\frac{N_iA_c(2\beta_m)\times10^{-8}}{l\Delta H_i/0.4\pi N_i}$$

$$=\frac{0.8\pi N_i^2A_c\beta_m\times10^{-8}}{l\Delta H_i}$$

$$K_i=2f(N_o/N_i)$$

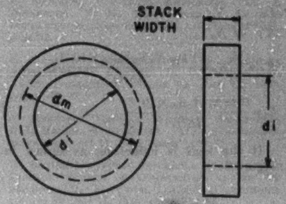


Figure 10. Toroidal core

Appendix II. Transfer Function of Simple Reactor Magnetic Amplifiers

Refer to Figure 7

$$K_i=\frac{\partial E_o}{\partial I_i}\qquad K_o=\frac{\partial E_o}{\partial I_o}$$

If $R_o=0$, then $E_o=0$ and $I_o=(N_o/N_i)I_i$.
Therefore $K_o=(N_o/N_i)K_i$.
From reference 7

$$I_o=I_i+\frac{E}{\omega L_{ac}}$$

where E = average voltage across reactor

$$\frac{\partial I_o}{\partial E_o}=\frac{1}{K_o}=\frac{1}{\omega L_{ac}}\frac{\partial E}{\partial E_o}=\frac{1}{\omega L_{ac}}$$

L_{ac} = a-c inductance of one unsaturated reactor

$$K_o=\omega L_{ac}$$

$$K_o=\frac{2\pi fN_o0.2A_c\beta_m\times10^{-8}}{l\Delta H_i/0.4\pi N_o}$$

$$K_o=\frac{1.6fN_o^2A_c\beta_m\times10^{-8}}{l\Delta H_i}$$

$$E_i=R_iL_i+p\frac{L_o}{K_i}E_o$$

$$E_o=K_i\left(I_i-\frac{N_o}{N_i}I_o\right)$$

$$E_o=\frac{1}{1+p\frac{L_i}{R_i}}\left(\frac{K_i}{R_i}E_o-\frac{N_o}{N_i}K_iI_o\right)$$

If the load is resistive, $I_o=E_o/R_o$

$$E_o=\frac{K_i}{R_i}\frac{1}{\left(1+\frac{N_oK_i}{N_iR_o}\right)}\frac{1}{1+p\frac{L_i}{R_i}}\frac{1}{1+\frac{N_oK_i}{N_iR_o}}E_i$$

if $\frac{N_oK_i}{N_iR_o}\gg1$

$$E_o=\frac{R_o}{R_i}\frac{N_i}{N_o}\frac{1}{1+P\frac{L_iN_iR_o}{R_iN_oK_i}}E_i$$

$$A_p=\frac{R_oN_i}{R_iN_o}\qquad T=\frac{L_iN_iR_o}{R_iN_oK_i}$$

$$T=\frac{1}{4f}\frac{R_o}{R_i}\left(\frac{N_i}{N_o}\right)^2$$

$$\frac{L_i}{K_i}=\frac{1}{4f}\frac{N_i}{N_o}$$

$$K_i=\frac{N_i}{N_o}K_o=\frac{1.6\pi f N_oN_iA_c\beta_m\times10^{-8}}{l\Delta H_i}$$

$$L_i=\frac{0.4\pi N_i^2A_c\beta_m\times10^{-8}}{l\Delta H_i}$$

If the load is inductive, as shown in Figure 8

$$I_o=\frac{E_o}{R_o\left(1+p\frac{L}{R_o}\right)}$$

$$E_o=\frac{1}{1+p\frac{L_1}{R_1}}\left(\frac{K_1}{R_1}E_1-\frac{N_o}{N_1}\frac{K_1}{R_o}\frac{E_o}{\left(1+p\frac{L}{R_o}\right)}\right)$$

$$E_o=\frac{K_i}{R_i}\frac{1}{1+\frac{N_oK_i}{R_oN_i}}\times\frac{1+pL/R_o\quad E_i}{1+p\frac{1}{1+\frac{N_oK_i}{N_iR_o}}\left(\frac{L_i}{R_i}+\frac{L}{R_o}\right)+p^2\frac{1}{1+\frac{N_oK_i}{N_iR_o}}\frac{L_iL}{R_iR_o}}$$

$$\frac{N_oK_i}{R_oN_i}>>1$$

$$E_o=\frac{R_o}{R_o}\frac{N_i}{N_o}\times\frac{(1+pL/R_o)E_i}{1+p\frac{N_iR_o}{N_oK_i}\left(\frac{L_i}{R_i}+\frac{L}{R_o}\right)+p^2\frac{N_iR_o}{N_oK_i}\frac{L_i}{R_i}\frac{L}{R_o}}$$

Appendix III. Definition of Terms

m=ratio of power amplification to time constant, per cycle or per second
A_p=ratio of change in amplifier output power to change in input power
T=amplifier 63 per cent response time, cycles or seconds
P_o=total amount of power dissipated in amplifier output circuit, watts
P_i=total amount of power dissipated in amplifier control circuit, watts
I_o=amplifier output or load current, average amperes
I_i=amplifier input or control current, average amperes
N_o=amplifier output winding turns per core
N_i=amplifier control winding turns per core
R_o=total resistance in amplifier output circuit, ohms
R_i=total resistance in amplifier input or control circuit, ohms
$\Delta P_o=(\Delta I_o)^2R_o$, watts
$\Delta P_i=(\Delta I_i)^2R_i$, watts
l=effective mean length of core magnetic path, centimeters
A_v=amplifier average voltage gain on a common turns base
A_I=amplifier average current gain on a common turns base
A_c=core cross-sectional area, square centimeters
A_w=core window area, circular mils

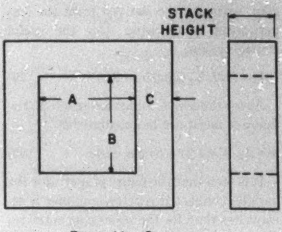


Figure 11. Square core

Appendix IV. Figure of Merit Related to Core Dimensions

Single-Phase Self-Saturating Magnetic Amplifier

$$m=2I_o\max N_o\frac{0.4\pi}{\Delta H_il}\text{ per cycle}\qquad(17)$$

It is now necessary to relate I_o max N_o to core dimensions. The maximum value of rms reactor current which essentially determines reactor heating at power frequencies is $I_o/\sqrt{2}\mu$. The maximum rms ampere-turns per reactor are then $I_oN_o/\sqrt{2}\mu$. Ampere-turns are directly related to copper and consequently to reactor window area.

For a given reactor volt-ampere rating the reactor load winding conductor can be assigned a minimum value of cross section per maximum rms ampere at a given supply frequency. This value ρ, the minimum permissible circular mils per rms ampere, is related to volt-ampere rating since reactors of a given volt-ampere rating at a given supply frequency have approximately the same heat dissipation capabilities.

$\rho I_oN_{o\,max}/\sqrt{2}\mu$ is equal to the reactor window area in circular mils times the load winding space factor.

$$\frac{\rho I_oN_{o\,max}}{\sqrt{2\mu}}=A_ws\text{ (circular mils)}$$

where s is the ratio between load winding copper area and total reactor window area.

Equation 16 now can be obtained from

$$I_oN_{o\,max}=\frac{\sqrt{2\mu}}{\rho}A_ws\text{ ampere turns}$$

$$m=2\sqrt{2\mu}\frac{A_ws}{\rho}\frac{0.4\pi}{\Delta H_il}\text{ per cycle}\qquad(18)$$

Appendix V. Figure of Merit Related to Core Dimensions

Toroidal and Square Cores Single-Phase Self-Saturating Magnetic Amplifiers

$$m=2\sqrt{2\mu}\frac{A_ws}{\rho}\frac{0.4\pi}{\Delta H_il}\text{ per cycle}\qquad(18)$$

Toroidal Core Construction (Figure 10)

$A_w=di^2\times10^6$ (centimeters)

where di is core inside diameter in inches.

$l=2.54\pi\ dm$ (centimeters)

where dm is core mean diameter in inches.

Then

$$m=2\sqrt{2\mu}\frac{sdi^2}{\rho}\times10^6\frac{0.157}{\Delta H_idm}\text{ per cycle}\qquad(18A)$$

Rectangular Core Construction (Figure 11)

$A_w=4/\pi ab\times10^6$ (circular mils)

where a and b are in inches.

$l=(2a+2b+4c)2.54=2(a+b+2c)2.54$

where a, b, and c are in inches.

Then

$$m=\frac{8\sqrt{2\mu}}{\pi}\frac{s}{\rho}ab\times10^6\times\left[\frac{0.08\pi}{\Delta HI(a+b+2c)}\right]\text{ per cycle}\qquad(18B)$$

For the special case of $a=b$ or a square window area

$$m=\frac{8\sqrt{2\mu}}{\pi}\frac{s}{\rho}\times10^6\left[\frac{0.04\pi}{\Delta H_i(a+c)}\right]\text{ per cycle}\qquad(18C)$$

Appendix VI. Experimental Check of Equations 14 and 18

$$m=2A_I\text{ per cycle}\qquad(14)$$

$$m=2\frac{\sqrt{2\mu}s}{\rho}A_w\frac{0.4\pi}{\Delta H_il}\text{ per cycle}\qquad(18)$$

Amplifier to be checked has the following characteristics:

Core material, Hipernik V
Core dimensions (toroidal)
  Outside diameter, 2$^1/_2$ inches
  Inside diameter, 1$^1/_2$ inches
  Strip width, $^1/_2$ inch
N_o, 2,300 turns number 26 American Wire Gauge

Two cores are connected in a single-phase self-saturating circuit giving these measured characteristics:

Rated supply voltage, 115 volt, 60 cycles per second
Rated power at the load, 30 watts
Rated load resistance, 125 ohms
Amplifier internal resistance, 60 ohms
Measured current gain on common turns base, 575
ΔH_i, 0.18 oersted
di, 1.5 inches
dm, 2 inches
Rated load current I_o, 0.56 ampere

Operating at $N_i^2/R_i=875$

$T=10$ cycles measured
$A_p=10{,}200$
$T_r=2.1$ cycles measured

From equation 9

$$m=\frac{A_p}{T}$$

$$=\frac{10{,}200}{10\text{-}2.1}=1{,}290\text{ per cycle}$$

From equation 14

$$m=2A_I$$

$m=2(575)=1{,}150$ per cycle

To check equation 18, ρ and s must be calculated from the available data.

Number 26 American Wire Gauge has 253 circular mils

$$\rho=\frac{253}{0.56}\sqrt{2(0.95)}=575 \text{ circular mils per rms ampere}$$

$$s=\frac{253\times 2300}{(1.5)^2\times 10^6}=0.26$$

$$m=2\sqrt{2\mu s}\,\frac{di^2}{\rho}\times 10^6\,\frac{0.157}{\Delta H_t dm} \text{ per cycle} \tag{18A}$$

$$m=2\sqrt{2(0.95)}(0.26)\,\frac{(1.5)^2}{575}\times 10^6\left(\frac{0.157}{0.18\,(2)}\right)$$

$=1{,}145$ per cycle

References

1. Self-Saturation in Magnetic Amplifiers, W. J. Dornhoefer. *AIEE Transactions*, volume 68, part II, 1949, pages 835–50.

2. Predetermination of Control Characteristics of Half-Wave Self-Saturated Magnetic Amplifiers, Henry Lehmann. *AIEE Transactions*, volume 70, part II, 1951, pages 2097–2103.

3. Determination of Steady-State Performance of Self-Saturating Magnetic Amplifiers, Edward J. Smith. *AIEE Transactions*, volume 69, part II, 1950, pages 1309–17.

4. Applying Magnetic Amplifiers, W. J. Dornhoefer, V. H. Krummenacher. *Electrical Manufacturing* (New York, N. Y.), September 1951, page 92.

5. General Characteristics of Magnetic Amplifiers, L. A. Finzi, D. C. Beaumariage. *AIEE Transactions*, volume 69, part II, 1950, pages 919–27.

6. Response Time of Magnetic Amplifiers, E. L. Harder, W. F. Horton. *AIEE Transactions*, volume 69, part II, 1950, pages 1130–41.

7. Series-Connected Saturable Reactor with Control Source of Comparatively Low Impedance, H. F. Storm. *AIEE Transactions*, volume 69, part II, 1950, pages 756–65.

8. Applying Magnetic Amplifiers, W. J. Dornhoefer, V. H. Krummenacher. *Electrical Manufacturing* (New York, N. Y.), April 1951, page 112.

No Discussion

Automatic Switching for Nation-wide Telephone Service

A. B. CLARK
FELLOW AIEE

H. S. OSBORNE
FELLOW AIEE

Synopsis: A plan for automatic long-distance switching, which ultimately will embrace the entire area of the United States and extend into Canada and perhaps Mexico has been formulated and important steps have been taken toward its realization. The plan contemplates that when a telephone customer places a call with a long-distance operator, this operator will be able to establish a connection to any desired telephone simply by playing a 10- or 11-digit code into an automatic mechanism. She will receive distinctive signals when the called telephone answers or when the telephone or the toll circuits are busy. Thus she will control completely the establishment of the connection and will have available to her the information necessary for proper billing of the call. The plan also contemplates that telephone customers ultimately will be able to dial long-distance calls themselves, wherever may be the locations of the calling and called telephones.

EVER SINCE the invention of the telephone 76 years ago, development work has been pressing forward both in telephone transmission and in switching. These two fields have been closely interrelated in the development of telephone service on a nation-wide basis, and neither could have progressed as it has without corresponding progress in the other.

The first development of equipment for the mechanical switching of telephone lines was the local dial system to enable one customer to be connected with another in the same town. It was a natural step to develop the equipment so that operators in nearby towns could complete toll calls through this local dial equipment. This was done first by using the local equipment and then with progressive modifications making it more and more suitable for toll.

By these means through the decades of the 1920's and 1930's regional networks were developed for operator toll dialing, using step-by-step types of equipment, particularly in Southern California, Connecticut, and Ohio. Also many short-haul toll calls in metropolitan areas were handled in connection with the panel-type dial equipment which was developed for automatic switching in these areas.

Also during this period the range of customer dialing in large metropolitan areas was extended, where local service is measured by message registers, through arrangements for the multiple registration of calls for which the charge was more than one local unit.

An important feature of switching development in this period was the perfecting of common control switching systems for large metropolitan areas endowed with a high degree of intelligence and great reliability.[1] As will be shown, still more extensive and complicated functions must be performed by the common control systems of a nationwide automatic switching system.

Also throughout this period great advance was made in the quality and stability of long-distance circuits. Telephone connections, some with as many as five circuits in tandem, were being established regularly by telephone operators with satisfactory over-all transmission. The limitation was in the speed and accuracy with which multiple switches could be made by operators rather than in the over-all transmission characteristics.

Several factors have worked together to bring about a big expansion of long-distance telephone service. These include the great growth in the number of telephones in service, improvements in long-distance transmission, in switching, and in methods of traffic operation. Since automatic switching becomes increasingly attractive as the traffic density increases, this large growth pointed toward the desirability of further mechanizing the switching operations.

In 1943 there was cut into service in Philadelphia the first installation of the number 4 toll crossbar system.[2] This system was designed to enable general automatic switching of toll connections in and out of large metropolitan areas and had many of the capabilities necessary for nationwide switching.

The various considerations already mentioned, coupled with the success of the number 4 installation at Philadelphia, led to studies of the service and operating results which might be expected from a nation-wide extension of automatic switching. The conclusion was reached that this would be a desir-

Paper 52-201, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted May 23, 1952; made available for printing May 23, 1952.

A. B. Clark is with the Bell Telephone Laboratories, Inc., New York, N. Y., and H. S. Osborne is with the American Telephone and Telegraph Company, New York, N. Y.

able objective of the Bell System companies and would result in a very substantial further improvement in the speed and accuracy of handling of long-distance messages. Accordingly, during the next few years, a national plan was prepared and was adopted by the telephone companies.

General Plan for Nation-wide Automatic Switching

The features of this nation-wide plan and the present status of its application form the subject of three companion papers.[3–5] The basic requirements to be met in the development of this plan included the following:

1. It should be suitable for the nation-wide extension of automatic switching both by originating toll operators and by the customers direct.

   When this work was commenced it was clear that a program leading toward general nation-wide operator dialing was desirable. Subsequent developments have confirmed the wisdom of making the basic plan consistent with general nation-wide customer dialing as well, since it now appears that a very wide extension of this form of service will become desirable.

2. The plan must provide for satisfactory over-all service between any two telephones in this country and Canada.

   Under manual operation satisfactory over-all service was provided for by the general toll switching plan in use since about 1930. This plan is modified to recognize the far greater speed and accuracy of automatic switching compared with manual switching. This involves also modifications of transmission design standards so that the over-all connections will continue to be satisfactory.

3. The system must be designed for instantaneous service, so that delays due to lack of circuits or equipment would be very infrequent. This is necessary, both from the standpoints of service and the avoidance of tieups, particularly of the automatic switching machinery.

   A trunking system must, therefore, be devised which will meet most economically this requirement, considering over-all costs of lines, switching equipment, and operation.

4. Machines must be designed for use at strategic points in the network, called control switching points, to perform automatically the various tasks required to make the over-all plan operative and economical.

5. The entire plan must be such as to provide satisfactorily for growth, for flexibility to meet changing conditions, and for minimum over-all costs of operation.

Fundamental Plans for Toll Plant

Mr. Pilliod's paper[3] discusses the fundamental layout of plant for nation-wide operator toll dialing. This is subject to changes from time to time with further specific studies, as is the case with all far-reaching fundamental plans of this type. The additional requirements imposed by nation-wide customer dialing are still under study as will be discussed a little later.

The national toll switching plan is modified so that there may be a maximum of eight toll circuits switched together to connect any two telephones compared with the previous limit of five.[6] In order to handle the entire traffic of the country, approximately 100 control switching points are necessary at which highly intelligent common control switching systems of the number 4 crossbar-type will be placed.

A very important feature of the layout is a trunking plan providing for a high degree of use of alternate routes. To design all of the toll circuit groups of the country for a no-delay service would be very expensive. However, taking advantage of the extreme rapidity of automatic switching and the ability to build into the machine capacity for using a large number of alternate routes, a trunking system has been devised in which only about one-sixth the toll circuit groups of the country need be engineered on a very liberal basis. These are called final groups and are the groups to which the machine ultimately appeals if all of the more direct circuit groups are busy. These more direct circuit groups can then be engineered on a basis providing for high usage of the circuits, recognizing that when one group is busy the machine appeals to another and so on until as a last resort the final group is used.

In determining means for handling all of the toll messages with a relatively small number of control switching points, tremendous advantage was derived from modern transmission developments, particularly carrier systems which give a great economy from the concentration on a long-distance route of large numbers of telephone circuits—numbers often running into the thousands. As a result, a considerable degree of circuitous routing and back hauling of circuits is economical if by these means the circuits can be concentrated on heavy routes. This in turn lends itself to a plan using a minimum of control switching points.

Nation-wide Numbering Plan

In the previous use of automatic switching by toll operators, the operators were furnished with codes by means of which could be selected the various circuits necessary to reach the destination. These codes were dialed, followed by the local number of the called party. With this system, toll operators calling a given telephone from different remote cities would, in general, use different codes corresponding to the different circuit groups which they must select.

For nation-wide toll dialing even by operators this system would have impossible complications, and for nation-wide customer dialing it is clear that the code to be dialed must represent uniquely the office which serves the called telephone and that office only and not be dependent upon the route to be followed to reach it. In other words, it involves the development of what is called a destination-type code. Another description of this code plan is to say that for toll dialing purposes each telephone in the country (and Canada) must have a distinctive telephone number different from that of every other telephone.

It is also clear that as a practical matter this number should be based upon the local telephone number of the customer prefixed by a minimum number of digits, following easily understood rules.

To bring this about has involved a very high order of planning. Such a plan has been perfected and currently forms the basis for the determination of the coding of all new telephone offices and for changes in office codes when these are necessary. The development of this is the subject of Mr. Nunn's paper.[4]

Customer Toll Dialing

When the customer is to dial long-distance calls directly without assistance from any operator, two additional requirements are imposed beyond those necessary for nation-wide operator dialing.

1. The customer normally is connected to a local central office, but for the purpose of nation-wide toll dialing he must be connected to the nation-wide toll network. At present he does this by dialing a code such as "211" which connects him with the long-distance operator. This procedure could be continued. However, since the customer must in any event dial 10 digits for the longest hauls to designate the called telephone, it is desirable if possible to cut out this preliminary step. That would mean modifying the local central office equipment so that it would receive the 10 digit numbers and transmit them on to the toll equipment. This is a simple undertaking for local central offices using the latest type of local central office equipment, called number 5 crossbar, which was designed with this in view.[7] For older types of equipment, the job is more difficult.

2. The switching equipment must be provided with automatic means for recording

all of the information necessary for charging the call. In the case of operator dialing this is now done manually by the operator.

Great advances have been made in recent years in the development of automatic message-recording equipment. In 1944 there was placed in service in California the first installation in this country of automatic ticketing equipment.[8] This equipment is associated with step-by-step local switching equipment and automatically prints for each call a ticket similar to that prepared by the operator with manual operation. In 1948 there was installed in Media, near Philadelphia, a greatly improved type of message recording equipment in which the information appears in the form of punched holes in a tape.[9] This equipment is much more economical than the earlier system and also lends itself to the automatic preparation of toll statements or bills.

The present forms of equipment have been designed to be associated with local central offices. A careful study has been made of their field of application and the basic plan necessary to provide for a general nation-wide extension of customer dialing. This indicates that there will be a large field for automatic message accounting equipment associated with the toll network and arranged to receive orders for toll messages from a number of local dial offices. This centralized Automatic Message Accounting equipment, as it is called, is under development and an initial installation will be made next year in Washington, D. C. In this installation the range of customer dialing will be limited and certain service features will be lacking, which it is planned to add later.

The nation-wide extension of customer toll dialing involves many operating problems in addition to those relating to the design of the plant. These problems involve the extent to which customers wish to dial long-distance calls, requiring 10 pulls of the dial, the accuracy of dialing, the treatment of wrong numbers, provision for giving subscribers information regarding telephone numbers in distant cities, information on charges, and many other questions.

Recognizing that the best way to develop these questions is a trial, arrangements were made to open such a trial last fall at Englewood, N. J. This office is equipped with a number 5 crossbar system so that arrangements for such a trial could be made readily there. The Englewood customers are able to dial directly any of about 11,000,000 telephones in 10 metropolitan areas scattered throughout the country, including Boston, New York, Pittsburgh, Cleveland, Chicago, and San Francisco and the Bay area.

The results of this trial have been very encouraging. Subscribers are continuing to dial over 95 per cent of all the calls which can be dialed. Errors due to wrong numbers are at a minimum and other difficulties are relatively low. In so far as this trial can answer the questions, the results are all in favor of the nation-wide extension of customer dialing as the development and installation of facilities suitable for this purpose make it possible to do so.

In view of the prospect of nation-wide customer dialing, fundamental plan studies are now being made by the Telephone Companies throughout the country of the whole layout of plant including the distribution of centralized automatic message accounting equipments with the future general application of this method of operation. The present indication is that the number of points at which toll-operating centers will be required will be greatly reduced. This will react in important ways on the design of telephone buildings, telephone equipment installations, and toll circuit routes.

Automatic Toll Switching and Accounting Equipment

All of these plans depend upon the successful development of striking innovations in toll switching and automatic message accounting equipments. The plans in turn react upon the features to be incorporated in such equipments and upon the schedule of their development. Mr. Shipley's paper[5] tells about the more important features of these equipments and the problems which are involved in their development.

Conclusions

Experience with operator toll dialing shows clearly that it provides a marked improvement in toll service. This improvement will increase as progress is made toward the full application of the nation-wide automatic switching plan.

The development of long-distance dialing by customers is at an early stage. The results of recent trials, however, indicate that nation-wide customer dialing has service advantages and generally will be received with enthusiasm by telephone users. It is anticipated, therefore, that customer dialing will expand rapidly both on a regional and also on a nation-wide basis.

The service advantages of nation-wide automatic switching are not measured entirely by the increased speed and improved accuracy of connections. An important factor is the continued ability of the telephone system to meet the rapidly increasing demand for telephone service without making excessive demands on the available supply of labor. The development of local dial operation was absolutely necessary to handle the great growth of local telephoning which has taken place. Today, in many places, requirements for people for toll operations are very heavy and an increased amount of automatic toll switching is becoming more and more necessary to make possible handling the rapidly increasing number of long-distance telephone messages.

With this development there has been a marked increase of employment. The Bell companies today employ 244,000 operators compared with 131,000 in 1941. They also have employed many people to build and install about $300,000,000 worth of toll dialing equipment, to construct places to house it, maintain it, and carry out operating rearrangements.

With respect to the future, even with the nation-wide automatic switching plan in full operation and the local central offices arranged to permit customer dialing, there still will be a large amount of work for operators. They will be required to handle information and assistance traffic, person-to-person calls, collect calls, and other classes of calls which do not lend themselves to customer handling, as well as any individual calls which the customers may not wish to dial themselves.

The Bell companies necessarily have taken the lead in planning and applying these new developments. The plans, however, are all laid in such a way as to include telephone users in independent telephone company offices. The independent companies are being kept fully informed of these plans as they develop and are participating, as the development of their own plant makes it practicable and desirable, in extending the benefits of the new forms of operation to their own customers.

This long-term development has required the very close co-operation of all parts of the Bell system—American Telephone and Telegraph Company General Department, Bell Telephone Laboratories, Western Electric Company, Long Lines, and all of the Bell operating companies. Each installation of equipment and circuits and each operation is a part of a nation-wide system and must be closely co-ordinated. The close interrelation and working together of the various parts of the Bell Telephone System, research and development, manufacturing, engineering, and operating are necessary for the effective planning and execution of this tremendous project.

References

1. **Crossbar Dial Telephone Switching System, F. J. Scudder, J. N. Reynolds.** *Electrical Engineering* (*AIEE Transactions*), volume 58, May 1939, pages 179–92.

2. **Crossbar Toll Switching System, L. G. Abraham, A. J. Busch, F. F. Shipley.** *Electrical Engineering* (*AIEE Transactions*), volume 63, June 1944, pages 302–09.

3. **Fundamental Plans for Toll Telephone Plant, J. J. Pilliod.** *AIEE Transactions*, volume 71, part I, September 1952, pages 284–56.

4. **Nation-wide Numbering Plan, W. H. Nunn.** *AIEE Transactions*, volume 71, part I, September 1952, pages 257–60.

5. **Automatic Toll Switching Systems, F. F. Shipley.** *AIEE Transactions*, volume 71, part I, September 1952, pages 261–69.

6. **A General Switching Plan for Telephone Toll Service, H. S. Osborne.** *AIEE Transactions*, volume 49, October 1930, pages 1549–57.

7. **The Number 5 Crossbar Dial Telephone Switching System, F. A. Korn, James G. Ferguson.** *AIEE Transactions*, volume 69, part I, 1950, pages 244–54.

8. **Automatic Ticketing of Telephone Calls, O. A. Friend.** *Electrical Engineering* (*AIEE Transactions*), volume 63, March 1944, pages 81–88.

9. **Fundamentals of the Automatic Telephone Message Accounting System, John Meszar.** *AIEE Transactions*, volume 69, part I, 1950, pages 255–69.

No Discussion

Fundamental Plans for Toll Telephone Plant

J. J. PILLIOD
FELLOW AIEE

THIS paper covers the general switching plan and fundamental plant layout proposed for handling telephone toll messages throughout the United States and Canada using automatic toll switching.

There has been rapid growth in the number of telephones and in the volume of toll traffic, particularly long haul. Toll facilities are provided under fundamental plans, an essential part of which is a toll switching plan for setting up connections quickly between any two telephones. The introduction of mechanical operation and the general improvement in the transmission performance of the communication plant over a period of years make the introduction of certain modifications in the fundamental plans possible and advantageous at this time. The important new features and the service improvements which are provided by the proposed plans are outlined in this paper. The principal types and characteristics of circuit facilities available for use in the intertoll network are also described.

General Aspects of Toll Switching Problems

Switching plans providing for the systematic routing of toll telephone traffic have been employed by the communication industry for many years. These plans have contributed directly to the high quality of long-distance telephone service enjoyed by the public in the United States and Canada. This generally excellent service is the result of the co-operative work of many organizations including the Bell operating companies, many independent connecting companies and others in the United States as well as in adjoining countries. The techniques employed today reflect a great amount of research and engineering and improvements in manufacturing skill and in construction, maintenance, and operating methods developed over a period of many years.

Throughout the United States and Canada there are approximately 20,000 different places—cities, towns, and villages—that serve as toll connecting points. The telephone offices in each of these places have access through the toll network to practically all of the 50,000,000 telephones in the United States and Canada and also to most of the telephones in the rest of the world. Currently, the Bell operating companies are handling toll calls at an average rate of over 7,000,000 during a business day. The many millions of different connection possibilities which this number of calls involves require a definite and comprehensive switching plan.

Whenever practicable and economical, direct circuits are used to handle toll message traffic between two given points. Much of the traffic in the country is handled this way. However, a substantial volume of business, about 20 per cent, is handled as a matter of economy, by switching toll circuits together. Although the volume of traffic between different points may vary over a wide range, it is nevertheless important that adequate service be provided for all possible connections. For example, there are about 110 circuits from Chicago terminating in the toll office serving Minneapolis and St. Paul. These handle about 5,500 calls per day. On the other hand, only a few calls a year may be involved between some point in Western Minnesota and a point in Florida. The switching plan described in this paper is devised for the purpose of efficiently and effectively establishing connections between any two points regardless of their separation and regardless of whether traffic volume be a few calls per year or many calls per hour.

Elements of the Problem

In order to illustrate the problem, a specific example may be useful. Figure 1 is a map of Wisconsin and Minnesota on which nearly 1,200 circles indicate points at which exchange facilities may be connected to the toll network. The extent of the coverage in this area is typical of that found throughout the country.

The 150-odd larger circles represent existing offices known as toll centers, that is, places where operators record toll calls and perform other operations necessary to establish toll connections. These places have switching arrangements of various types depending on how they fit into the switching plan. Some may operate as control switching points in the nationwide plan as described later.

More than 1,000 smaller circles on the map represent tributaries, that is, towns where little or no toll operating is done. Toll connections to and from these points are completed at the toll centers which in general do the toll operating required.

Paper 52-202, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted May 23, 1952; made available for printing May 23, 1952.

J. J. Pilliod is Assistant Chief Engineer of the American Telephone and Telegraph Company, New York, N. Y.

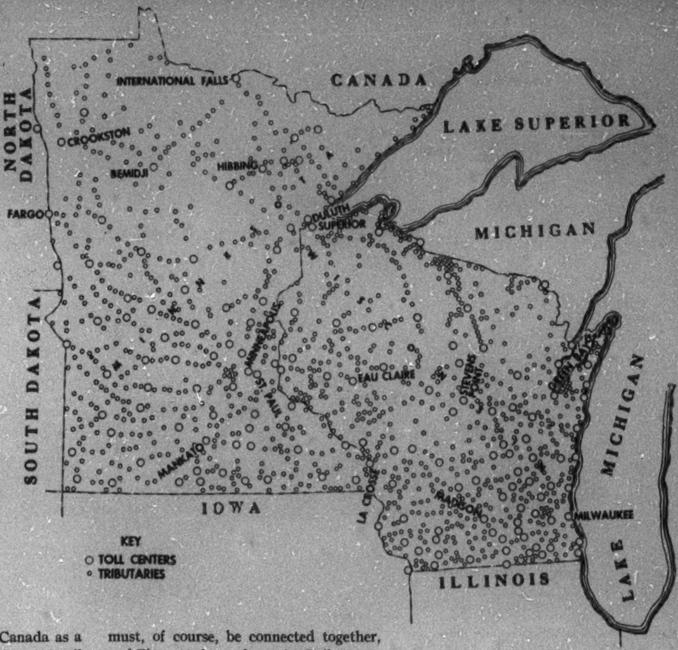


Figure 1. Toll centers and tributaries in Minnesota and Wisconsin

In the United States and Canada as a whole, there are approximately 2,600 toll centers. The remainder of the toll connecting points—about 17,500—are tributaries.

Figure 2 gives an idea of the variety and complexity of the network of circuit groups required to interconnect the toll centers in one area. Here each line represents a group of circuits known as intertoll trunks between two toll centers. Each group may contain anywhere from one to several dozen trunks. The location of the lines on the map is unrelated to the geographical routing of the trunks, and only a part of the circuit groups are shown. To get a complete picture, one should visualize that a cluster of relatively short circuit groups radiates from each toll center to its tributaries, of which there may be up to 15 or more.

Physically, the plant consists of a network of open wire lines, cables, and radio systems. On these, voice frequency or carrier operation is employed in each section, as required, to provide the necessary intertoll trunks. The routes of the lines in Minnesota and Wisconsin are shown by Figure 3. In this area there are no radio routes carrying telephone circuits, but a radio system between Chicago and Minneapolis is in the planning stage.

Areas like Wisconsin and Minnesota must, of course, be connected together, and Figure 4 shows the major Bell system toll routes that accomplish this. On a map of this kind it is not possible to include anything like the detail shown in Figure 3. One must visualize, therefore, that each state contains a network of routes generally comparable to those shown for Wisconsin and Minnesota.

This then represents the interconnection problem to be met by an orderly switching plan that will provide efficient, reliable, and fast toll telephone service between any two points.

Earlier Toll Switching Plans

Very early in the telephone industry it became evident that:

1. There must be a plan for connecting circuits together.
2. Switching centers with suitable equipment must be established in accordance with this plan.
3. Trunks must be provided in adequate numbers to connect every place to one or more switching centers and to interconnect the switching centers.
4. All this must be done in a way that makes it possible to provide good service at reasonable cost.

As time went on, early plans crystalized into what became known as the General Toll Switching Plan. A paper by Dr. H. S. Osborne outlined the principles of this comprehensive plan for handling telephone toll traffic in the United States and Eastern Canada.[1] It involved two classes of major switching centers, regional centers and primary outlets, and some classes of less important centers. It also set up methods of designing toll trunks to give adequate transmission efficiency on all possible toll connections. In use for the last 2 decades, this basic plan has been of great value in accommodating the tremendous growth of telephone toll business during this period.

Switching Plan for Nation-wide Toll Dialing

The earlier general switching plan was based on manual switching and on a toll plant made up, for the most part, of voice frequency circuits. The probability of operating irregularities and delays increases with the number of manual switches in tandem. Likewise, the transmission problem of operating many voice frequency trunks in tandem was so formidable that the number of intertoll

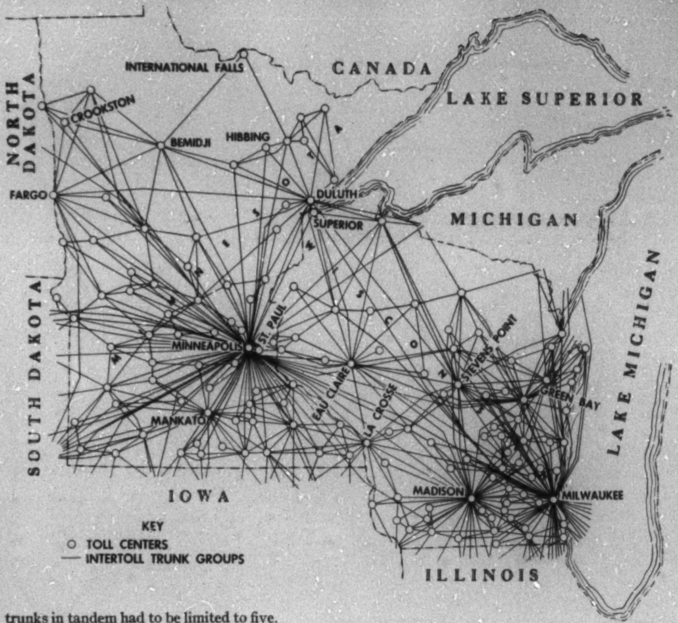


Figure 2. Principal intertoll trunk groups in Minnesota and Wisconsin

trunks in tandem had to be limited to five. In practice, switching was avoided wherever practicable and economical.

Impact of Mechanization and Improved Transmission Facilities

On the other hand, mechanical switching is very fast and is designed to be practically free of operating irregularities. Delays can be minimized by fast switching to alternate routes. Also, in the last 2 decades the use of carrier has grown from a relatively minor place in the toll plant to the point where it is now commonplace.[2] Carrier provides superior transmission performance. Limitations on switching are thus greatly reduced and economies are achieved under many conditions.

In addition, mechanization of local switching systems has proceeded rapidly. With mechanized toll switching, it is becoming possible to establish many toll connections with only a single toll operator and in some cases by customer dialing, without the assistance of any operator.[3,4]

Along with these developments has come tremendous growth in traffic. Since 1930 toll messages in the Bell operating companies and the Bell Telephone Company of Canada have more than trebled, growing from an annual volume of about 650,000,000 to about 2,000,000,000. Intertoll trunks over 25 miles in length have increased in number from about 28,000 to about 100,000. This continuing growth in traffic volume has required a large-scale development of plant facilities and has permitted a more extensive use of carrier than would have been practicable with a slower rate of growth.

Consideration of these factors which offer an opportunity to improve service has led to the gradual reorientation of the fundamental plans for the intertoll trunk plant which is now under way.

The New General Toll Switching Plan

Mechanization of switching and the use of improved transmission instrumentalities permits the design of the switching plant to be controlled primarily by the balance between the costs of transmission facilities and of switching facilities.

The new general toll switching plan contemplates as many as eight intertoll trunks in tandem on the most complex connections to be established. These eight trunks can be interconnected at switching points as described later. The plan further contemplates that wherever possible, the traffic will by-pass intermediate switching points. The number of switches that can be avoided depends on the volume of traffic between the two points concerned and on the traffic load at the time the connection is established.

The proposed plan provides a systematic grouping of switching points. Under this arrangement, each ordinary toll center (TC) serves a cluster of near-by-tributary points and has trunks to a home primary outlet (PO) which serves a cluster of toll centers. In some cases it appears practicable to utilize a simplified switching system at a PO, and in order to distinguish this type of center it has been designated a tandem outlet (TO). In turn, each PO or TO has trunks to a home sectional center (SC) which serves a section of the country varying in size from part of a state to all of several states, depending on the density of the population. Similarly, the United States and Canada are divided into nine regions, each having a regional center (RC) serving as a central switching point for all sectional centers in the region. One of these RC's (St. Louis) is termed the national center (NC). All of the higher orders of switching centers also act in the capacity of each of the lower centers. For example, any specific SC also acts as a PO and as a TC.

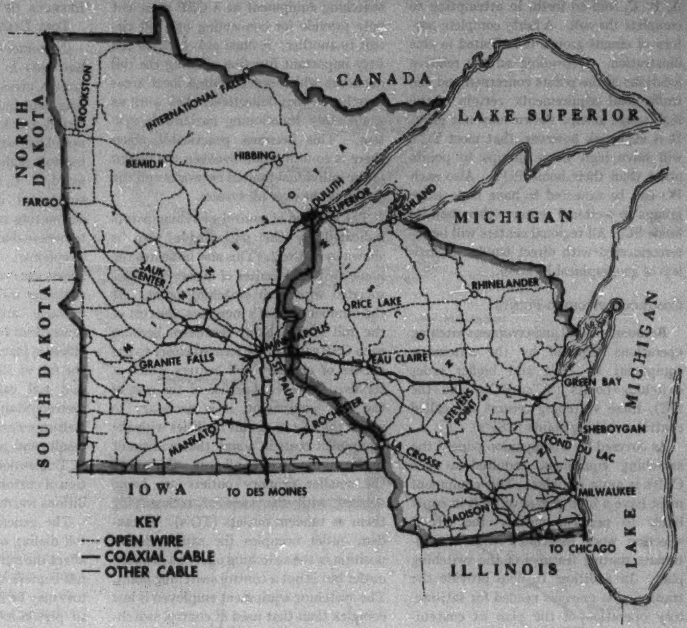


Figure 3. Bell System telephone toll routes in Minnesota and Wisconsin

This arrangement is illustrated in Figure 5 which covers approximately the same area as Figure 1 portraying the toll connecting points. Hibbing, Minnesota is shown as a representative toll center with the tributaries it serves. It is in the service area of the Duluth tandem outlet, the approximate boundaries of which are indicated. Duluth lies in the Minneapolis section, which includes a large portion of Minnesota, and is in turn one segment of the Chicago region which serves a little larger area than shown by Figure 5.

Under this arrangement, toll calls between two tributaries in the Hibbing toll center area can be completed by switching at the toll center. In a similar manner, any two points within the Duluth tandem outlet area can be served by switching at Duluth. The same treatment also applies for connections between any two points in the same sectional center area or in the same regional center area. For example, a connection from Hibbing to any point within the Chicago region (which involves more than six states as shown in Figure 7) requires no more intertoll links than Hibbing to Duluth, Duluth to Minneapolis, and Minneapolis to Chicago, and a corresponding number of links on through to another sectional center, and primary or tandem outlet to the toll center destination. Circuits between the toll center and tributaries are not referred to as intertoll trunks or lines but are classed as toll connecting trunks.

Where the volume of traffic warrants, direct circuits may be provided to by-pass the intermediate switching points included in the preceding example. Once such direct-circuit groups have been established, it is economical and advantageous from a switching standpoint to take advantage of their existence, using routes that involve a minimum number of switches. The basic routing plan is used when the more direct-circuit combinations are busy.

These routing arrangements contemplate the application of "high usage" and "final" trunk groups as an integral part of the plan. The high usage groups are direct groups which by-pass the higher order switching points wherever the routing of the call permits. These high usage groups can be engineered to carry high loads per circuit, with an adequate number of circuits in the final groups to take care of practically all overflows from the high usage groups during the heavy traffic periods. The high usage and final groups which could be used for routing calls between Hibbing, Minnesota and Davenport, Iowa are shown by Figure 5.

GENERALIZATION OF THE TOLL SWITCHING PLAN

The generalization of the arrangements discussed for the Chicago region is illustrated in Figure 6. This shows diagrammatically all types of switching points in two regions and also indicates the relative position occupied by the national center in the switching plan. On this chart, the solid lines represent the final groups of trunks and the dotted lines represent high usage trunks. Examination of this chart will indicate that the mechanical switching system need perform only relatively simple toll switching operations at the toll centers. At other points the system must attempt to complete the call over the most favorable routes, in planned sequence, until the final route is selected.

For example, from a given primary outlet such as PO1 on a call destined for a toll center in the other region such as TC2, the switching equipment would attempt to complete the call in sequence over the routes marked 1 to 6.

Should Route 6, which is the final route, be selected because all of the trunks in the high usage groups marked 1 to 5 were busy at the time, the switching equipment at the SC would in turn try routes marked

A, B, C, and so forth, in attempting to complete the call. A fairly complete pattern of circuit groups is indicated in this illustration. Depending on the relative locations of the points concerned and the traffic load requirements, certain of the high usage groups shown may not exist. It is expected, however, that most TC's will have high usage groups to points other than their home PO's. Also each PO can be expected to have high usage groups to sectional centers other than its home SC. All regional centers will be interconnected with direct trunks, regardless of geographical location.

CONTROL SWITCHING POINTS

Because of rapid and complex switching operations required by the automatic equipment at PO's and higher order switching points (SC's, RC's, and the NC), these switching centers are called control switching points (CSP's).

As covered by a companion paper,[6] the switching equipment required at the CSP's is quite complex. This equipment must have a high degree of built in capability to perform quickly the circuit selection work associated with the alternate routing features of the switching plan. In addition, to help provide the transmission margins needed for satisfactory operation of the plan as contemplated, it must be arranged to connect circuits on a 4-wire basis rather than on a 2-wire basis, the latter being the arrangement used at most toll centers. The switching equipment at a CSP must not only provide for connecting one toll circuit to another; it must also perform the very important function of tying the toll networks which serve limited local areas together so that collectively they work as a smoothly functioning nationwide system. This becomes practicable when there is co-ordination between the design of the individual limited networks and the design of the over-all system.

The location of control switching points indicated by the nation-wide plan is shown in Figure 7. This also indicates the home switching center of higher order associated with each switching point. As the number of CSP's increases, the cost of the toll circuit plant decreases because each CSP can then be located closer to the cluster of ordinary toll centers which it serves. However, because of the cost of the CSP equipment, it is necessary to weigh the cost of circuit facilities with the equipment costs in a way that will result in the minimum over-all cost. Certain of the smaller primary outlets are being studied with the view of reclassifying them as tandem outlets (TO's). A tandem outlet occupies the same relative position in the switching plan as a primary outlet but is not a control switching point. The switching equipment employed is less complex than that used at control switching points and, therefore, provides for only limited alternate routing and does not have all of the advantages of 4-wire transmission.

EFFECTS OF CUSTOMER AND OPERATOR TOLL DIALING

Customer dialing of short-haul toll calls has been in use, particularly in metropolitan areas, for some years. A trial of long-haul customer dialing over the intertoll trunk network and through the switching equipment provided for operator toll dialing was instituted at Englewood, N. J., in the fall of 1951. The local equipment includes automatic message accounting and permits Englewood customers to dial directly to about 11,000,000 telephones in ten metropolitan areas across the country. A trial installation of customer toll dialing, utilizing automatic message accounting equipment on a centralized basis rather than at each local office, is planned for Washington, D.C. in the fall of 1953. Initially customers will dial toll calls within the Washington metropolitan area and to such points as Baltimore and Annapolis. The favorable results and general acceptance of the trial at Englewood indicate extensive application of customer dialing of toll calls as conditions warrant.

The general introduction of customer toll dialing as this becomes desirable will affect the number and location of ordinary toll centers since calls handled by operators may be limited to assistance calls and to person-to-person, collect, and others which cannot be customer dialed. Indications are that toll operation for a number of smaller centers can be combined as the local service is converted to

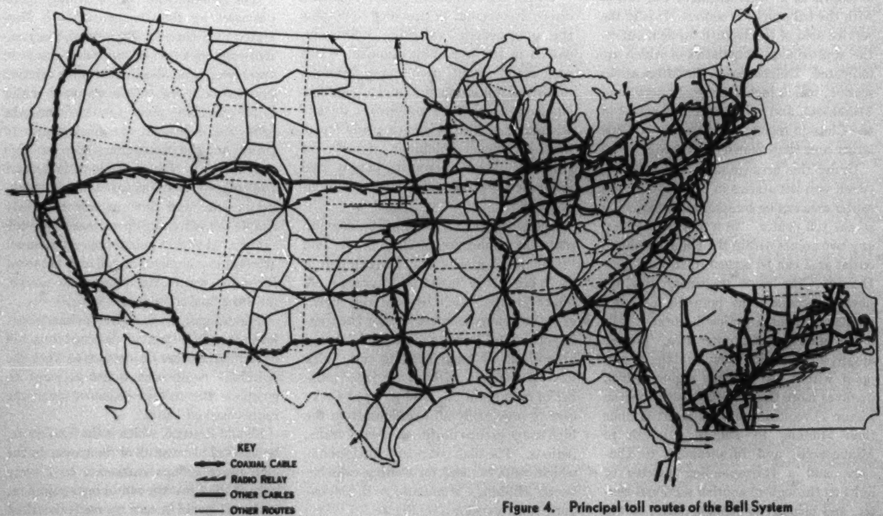
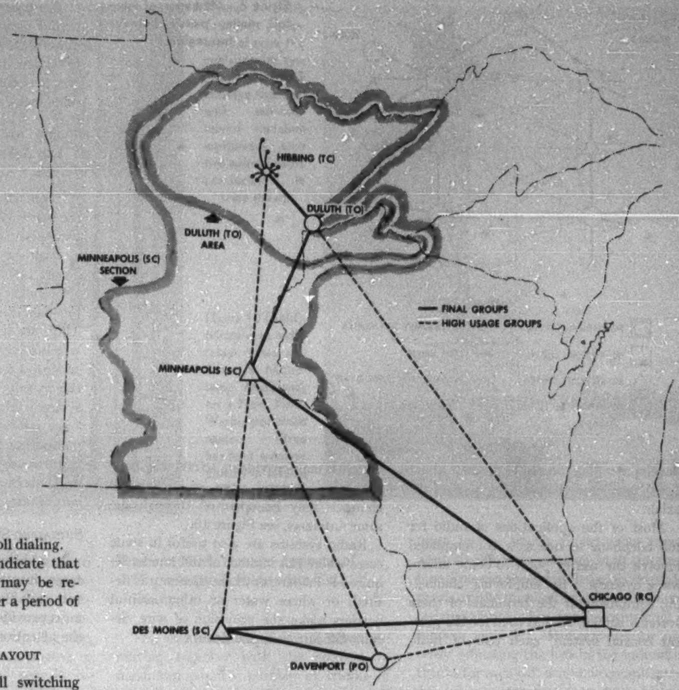


Figure 4. Principal toll routes of the Bell System

Figure 5. Intertoll trunks between Davenport, Iowa, and Hibbing, Minnesota, showing alternate routing possibilities



dial operation with operator toll dialing.

Studies now in progress indicate that the number of toll centers may be reduced by one-half or more over a period of years in many areas.

REACTIONS ON TOLL PLANT LAYOUT

The expanded general toll switching plan for nation-wide dialing, using a nation-wide numbering plan,[6] contemplates a degree of alternate routing far in excess of that used with the former switching plan designed for manual operation. This change along with the reduction in toll centers will have a marked effect on the normal flow of many traffic items through the intertoll network. As a result the arrangement of the present intertoll trunks will be significantly modified both in number, routing, and terminating points. It is necessary to take these facts into account in engineering toll plant additions so that they will lead toward an advantageous layout for future nation-wide dialing as well as meet the needs of the more immediate future. Fortunately, the effect is in the direction of greater concentration of circuits in main routes so that with the new cable and radio facilities available, over-all economy and better service should result.

Types of Transmission Facilities Used and Included in Switching Plan

The domestic toll network is an outgrowth of the demands of the business and the advance in communication technique over many years. At present, about 100,000 intertoll trunks over 25 miles in length and many thousand shorter toll trunks are in service throughout the country. They are provided generally by voice frequency or carrier frequency facilities. The choice of transmission facility on a given route is dependent on a number of factors, such as cost, length of haul, number of trunks in the cross section, numbers of trunks to be terminated at intermediate points, the types of terrain to be traversed, storm and other conditions affecting service continuity, and the transmission requirements of the circuits to be provided.

Voice frequency facilities equipped with repeaters as required are used on both open-wire lines and cables. At voice frequencies it is customary to derive three trunks known as a phantom group, from two pairs of open wires or from one "quad" (2 pairs) of loaded cable conductors. In general, the use of voice frequency facilities is now limited to shorter circuits.

Considerations of economy and service improvement led to the introduction of carrier operation into all types of toll plant as rapidly as the state of the art permitted. This directly affects the toll switching plan from the standpoint of routing and location of switching centers.

At present, carrier systems use four broad categories of facilities: open wire, conventional paired or quadded cables, coaxial cable, and radio.

Several types of open-wire carrier systems permitting from 1 to 15 telephone channels above the frequency band of the voice channel are now in use. In general, these systems are used where trunk cross sections are relatively small and where the terrain and weather conditions make open-wire lines economical.

Cable carrier systems presently permit the operation of up to 12 telephone channels on two pairs of cable conductors. These conductors may be in one cable or divided between two separate cables, depending on the type of carrier system, see Figure 8. Coaxial cable transmission systems currently provide up to 600 telephone channels per pair of coaxials, see Figure 9. A new coaxial system under

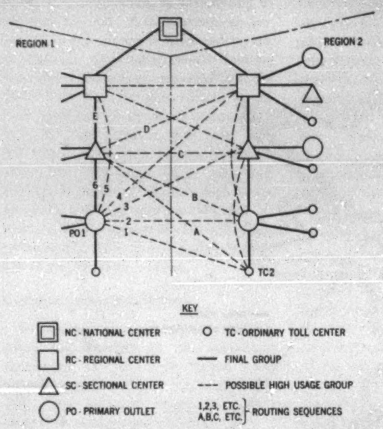

REGION 1
REGION 2
E
D
C
B
A
6
5
4
3
2
1
PO1
TC2
KEY
NC - NATIONAL CENTER
RC - REGIONAL CENTER
SC - SECTIONAL CENTER
PO - PRIMARY OUTLET
TC - ORDINARY TOLL CENTER
FINAL GROUP
POSSIBLE HIGH USAGE GROUP
1,2,3, ETC.
A,B,C, ETC.
ROUTING SEQUENCES

Figure 6. Illustration of intertoll routing pattern between two regions

development is expected to produce about 1,800 telephone channels per pair of coaxials.

Most of the applications of radio for toll telephone service now contemplated involve the use of point-to-point microwave systems. By employing channeling equipment at the terminals of these systems similar to that used for the present coaxial system, each pair of radio channels may provide up to 600 telephone channels. Several pairs of such radio channels may be operated through the same antennas, see Figure 10.

Radio systems are also useful in some cases where the number of toll trunks required is moderate, where diversity is desired, or where water or other natural barriers make the provision of wire circuits difficult or impracticable.

The type of facility to be used on a particular route is sometimes affected by requirements for other services such as teletypewriter, television network facilities, program facilities, private lines, and other factors.

TREND TO CARRIER-TYPE FACILITIES AND ADVANTAGES TO TOLL SWITCHING PLAN

About 70 per cent of the long-haul toll message mileage in Bell operating companies is provided on carrier-type facilities as contrasted with 7 per cent in 1930, see Figure 11.

From the transmission standpoint carrier facilities offer marked advantages. They are inherently of the 4-wire type which minimizes the number of possible singing and echo paths on a circuit. Also, the speeds of propagation over carrier systems are generally higher than over voice frequency systems thereby further minimizing the echo problem. These features are of great advantage in reducing limitations on circuit design and layouts of the general toll switching plan.

SIGNALING SYSTEMS

In addition to the ability to carry messages, intertoll trunks must be provided with suitable signaling facilities.[7,8] These must provide a means of: first, attracting the attention of the distant point, either

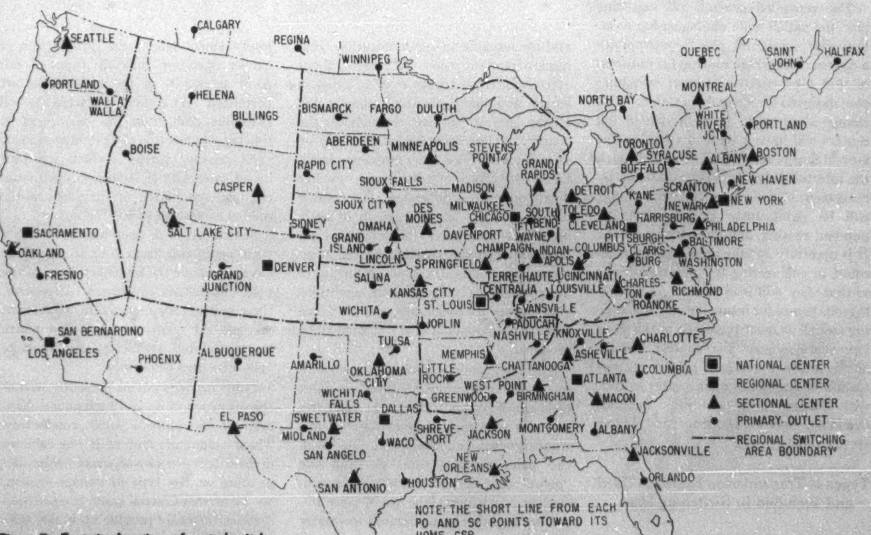

SEATTLE
CALGARY
REGINA
WINNIPEG
QUEBEC
SAINT JOHN
HALIFAX
PORTLAND
WALLA WALLA
HELENA
BILLINGS
BISMARCK
FARGO
DULUTH
NORTH BAY
MONTREAL
WHITE RIVER JCT
PORTLAND
BOISE
ABERDEEN
MINNEAPOLIS
STEVENS POINT
GRAND RAPIDS
TORONTO
SYRACUSE
ALBANY
BOSTON
RAPID CITY
BUFFALO
NEW HAVEN
CASPER
SIOUX FALLS
MADISON
DETROIT
KANE
SCRANTON
NEW YORK
SIOUX CITY
DES MOINES
MILWAUKEE
SOUTH BEND
TOLEDO
NEWARK
HARRISBURG
PHILADELPHIA
SACRAMENTO
SALT LAKE CITY
SIDNEY
OMAHA
CHICAGO
FT. WAYNE
CLEVELAND
PITTSBURGH
BALTIMORE
OAKLAND
GRAND ISLAND
DAVENPORT
CHAMPAIGN
INDIANAPOLIS
COLUMBUS
CLARKSBURG
WASHINGTON
FRESNO
GRAND JUNCTION
DENVER
LINCOLN
SPRINGFIELD
SALINA
TERRE HAUTE
CINCINNATI
CHARLESTON
RICHMOND
KANSAS CITY
CENTRALIA
LOUISVILLE
ROANOKE
ST. LOUIS
EVANSVILLE
WICHITA
PADUCAH
SAN BERNARDINO
JOPLIN
NASHVILLE
KNOXVILLE
CHARLOTTE
LOS ANGELES
PHOENIX
ALBUQUERQUE
TULSA
MEMPHIS
ASHEVILLE
AMARILLO
OKLAHOMA CITY
LITTLE ROCK
CHATTANOOGA
COLUMBIA
WEST POINT
ATLANTA
WICHITA FALLS
GREENWOOD
BIRMINGHAM
MACON
DALLAS
EL PASO
SWEETWATER
SHREVEPORT
JACKSON
MONTGOMERY
ALBANY
MIDLAND
WACO
SAN ANGELO
JACKSONVILLE
NEW ORLEANS
SAN ANTONIO
HOUSTON
ORLANDO
NATIONAL CENTER
REGIONAL CENTER
SECTIONAL CENTER
PRIMARY OUTLET
REGIONAL SWITCHING AREA BOUNDARY
NOTE: THE SHORT LINE FROM EACH PO AND SC POINTS TOWARD ITS HOME CSP.

Figure 7. Tentative locations of control switching points in the United States and Canada

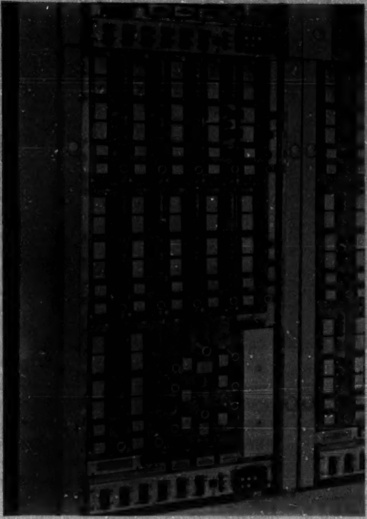


Figure 8 (left). Terminal equipment of type N1 cable carrier system. Provides 12 message channels with self-contained signaling equipment over two pairs of cable conductors in same sheath

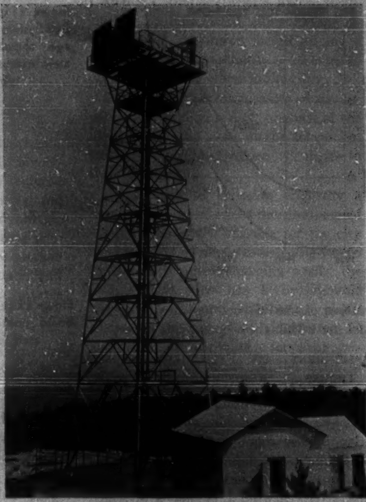


Figure 10 (right). Microwave radio relay tower at Cotoctin Mountain, Maryland, on a New York-Washington radio route. Three hundred message circuits in service with more planned

an operator or automatic equipment, to the fact that a connection is to be established; second, in the case of dial operation, transmitting coded information in the form of pulses for establishing the connection; and third, transmitting a general class of supervisory signals including connect and disconnect signals, on and off switch hook signals, recall signals, and busy signals which are essential to the efficient operation of the switching plant. The circuit design contemplated in the over-all plan must take into account this requirement for transmitting signals as well as speech, to obtain accuracy and speed in setting up and taking down connections.

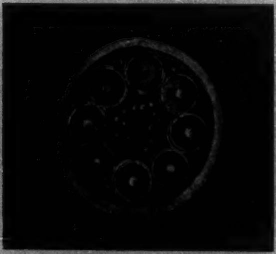


Figure 9. Coaxial cable. Cross section of cable containing four pairs of coaxials. Each pair can accommodate one 2-way coaxial carrier system

Transmission Design Aspects of Circuits for Nation-wide Toll Dialing

The more extensive use of alternate routing, together with the increase in maximum possible number of trunks in tandem associated with nation-wide toll dialing, tends to increase the problems of assuring adequate transmission of speech and signals on all possible connections. On the other hand, the use of 4-wire switching at important points and the definiteness of the routing patterns permit more effective use of the available facilities and thus tend to simplify the problem. Extensive studies indicate that, on the whole, the new toll switching plan will make feasible still further improvements in transmission. This is, of course, a desirable objective.

TRANSMISSION DESIGN OF TRUNKS

With dial operation, the number of trunks in tandem in a given toll connection may vary on successive calls. To avoid undesirable transmission contrasts and other adverse effects, it is important that every trunk be designed to operate as closely as possible to the theoretically correct transmission loss. The problem is complicated by the fact that the extent to which the echo, noise, and crosstalk will limit the performance of an individual link is not directly proportional to the length of the circuit. In fact, the minimum loss at which a particular circuit used singly or in various built-up combinations can theoretically be operated depends on the number, length, and characteristics of the other circuits connected in tandem with it. Arrangements for precisely adjusting the loss in the individual trunks for each call would be complicated. Adequate performance can be achieved, however, by compromise methods which provide for automatic adjustments in the loss of each trunk in accordance with:

1. When a trunk is switched to other intertoll trunks at both ends, it is operated at the minimum loss practicable. This loss is known as via net loss (VNL).
2. When the trunk is switched to another intertoll trunk at one end only, the loss is increased 2 decibels.
3. When the trunk is not switched to another intertoll trunk at either end a further loss of 2 decibels is added. This loss which is 4 decibels greater than the via net loss is known as terminal net loss (TNL).

The data and methods used in the der-

Table I. Comparison of Transmission Losses for Typical Intertoll Connections Realized under Manual Plan and Objectives under Dialing Plan

Number of Links in Intertoll Connection	Approximate Range of Losses Between Toll Centers in Decibels	
	Manual Plan	Dial Plan
1	4–12	4–8
2	8–14	5–12
5	9–20	6–13
8		7–13

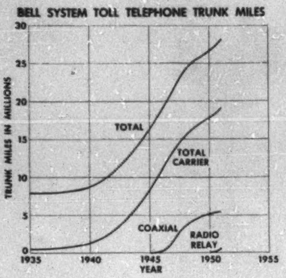


Figure 11 (left). Growth in Bell System intertoll trunk mileage showing trend toward more extensive use of carrier-type facilities

Figure 12 (right). Toll switchboard position with key set used for toll dialing

ivation of the via net loss are complex and not within the scope of this paper.

Assignment of Facilities Among Trunks

The definite routing patterns established for the toll machine switching operation impose more severe transmission conditions on certain classes of circuits than on others. For example, a trunk in a final group between a TC and a PO can become involved in an 8-link connection, whereas a trunk in a high usage group, say, between a PO and another PO will not be involved in more than a 3-link connection.

This creates a need and provides an opportunity for allocation of the available facilities among the various trunk groups in a way that will provide the best over-all service. For example, to the extent practicable it is desirable to assign carrier grade facilities to trunks in final groups that may be involved in connections with the maximum number of links. Facilities with less favorable transmission characteristics may then be reserved for trunks in groups that are used for connections involving fewer links.

Transmission Performance

Table I shows the approximate range of transmission losses between toll centers under the manual plan compared to ranges that appear practicable under the proposed fundamental plan, which, of course, permits more links in tandem.

Trunk Transmission Stability

It is as important that the transmission loss of a trunk used in the contemplated toll dialing network be maintained at or close to its assigned value at all times as that the assigned value be right. On multi-switched connections even a relatively small consistent excess or deficiency in the loss in the individual trunks can accumulate to over-all excesses or deficiencies in loss large enough to cause difficulty by making it hard for people to hear if the attenuation becomes too great or by creating excessive echo, crosstalk, or noise if the loss becomes appreciably less than normal.

This subject has been extensively studied for the past several years and it appears that some changes in practices and the introduction of new methods of measuring results will lead to marked improvements. It is of some interest that one of the major factors in securing improvement appears to be the application of a statistical method of evaluating performance along somewhat the same lines as the quality control methods used in other fields of industry.

Since, with operator toll dialing only one operator is involved in many connections and with customer toll dialing there is no operator on the connection, it is extremely important that everything be right. This is typical of the requirements of any large scale push-button operation, see Figure 12.

Conclusion

The fundamental plans proposed for telephone toll switching provide a basis for the progressive mechanization of toll service. The installation of suitable switching mechanisms at control switching points and the provision of toll trunks utilizing the new instrumentalities will implement the toll switching plan. The plan is sufficiently flexible to adjust for changes in the telephone art as they develop. Also, the plan can fit in with the requirements of those companies whose plants connect with the Bell operating network should they desire to arrange for operator or customer toll dialing.[9–11]

Average speed of service will be improved. The flexibility in plant design inherent in the new toll switching plan will increase service security and improve the utilization of the entire toll plant. In addition, adequate provision is made for the progressive introduction of customer toll dialing as this becomes practicable and desirable.

References

1. The General Switching Plan for Telephone Toll Service, **H. S. Osborne.** *AIEE Transactions*, volume 49, October 1930, pages 1549–57.

2. Carrier is King, **C. M. Mapes.** *Bell Telephone Magazine* (New York, N. Y.), volume XXVIII, Winter 1949–50, pages 191–203.

3. Operator Toll Dialing, **J. J. Pilliod, H. L. Ryan.** *Bell Telephone Magazine* (New York, N. Y.), volume XXIV, Summer 1945, pages 101–115.

4. Toll Dialing is Expanding Throughout the Nation, **E. W. Baker.** *Bell Telephone Magazine* (New York, N. Y.), volume XXX, Winter 1951–52, pages 253–64.

5. Automatic Toll Switching Systems, **F. F. Shipley.** *AIEE Transactions*, volume 71, part I, September 1952, pages 261–69.

6. Nation-wide Numbering Plan, **W. H. Nunn.** *AIEE Transactions*, volume 71, part I, September 1952, pages 257–60.

7. Application of Multifrequency Pulsing in Switching, **C. A. Dahlbom, A. W. Horton, Jr., D. L. Moody.** *AIEE Transactions*, volume 68, part I, 1949, pages 392–96.

8. Single-Frequency Signaling System for Supervision and Dialing over Long-Distance Telephone Trunks, **N. A. Newell, A. Weaver.** *AIEE Transactions*, volume 70, part I, 1951, pages 489–95.

9. Nationwide Operator Toll Dialing, **American Telephone and Telegraph Company.** *Telephony* (Chicago, Ill.), January 12, 1946.

10. New Toll Switching Plan for Nationwide Dialing, **American Telephone and Telegraph Company.** *Telephony* (Chicago, Ill), May 10 and 17, 1947.

11. Nationwide Toll Dialing—Use of Tandem CDO's, **American Telephone and Telegraph Company.** *Telephony* (Chicago, Ill). July 3, 1948.

12. Automatic Switching for Nation-wide Telephone Service, **A. B. Clark, H. S. Osborne.** *AIEE Transactions*, volume 71, part I, September 1952, pages 245–48.

No Discussion

Nation-wide Numbering Plan

W. H. NUNN
NONMEMBER AIEE

IN TELEPHONE language a numbering plan is exactly what the name implies, a plan or system of giving each telephone in a city, a town, or any geographical area an identity or designation which is different from that given every other telephone in this same area. This designation is the telephone number; it appears in the directory and in most cities on the telephone instrument itself. It is the address of the telephone in the telephone network. Just as it is essential for efficient postal and delivery service to have streets and house numbers clearly marked, it is important for good telephone service that the telephone numbering plan be such that it will be used with convenience and accuracy by the telephone customer.

A telephone number is comprised of two elements, a designation for the central office to which the telephone is connected and a number within the central office which identifies one particular telephone from all others served by that office. If there is only one central office in the city or town, the office designation is frequently omitted. A dial office is designed to serve up to 10,000 numbers with a limitation of four digits. Typical numbers are therefore MAin 2-1234, ADams-2345, 5-6789, and 3456, the office designations being MAin 2, ADams, and 5 with the last four digits in all cases representing the number within the central office.

There is a wide variation in the types of numbering arrangements in use today in the Bell System. This diversity arises from the fact that telephone communities vary greatly with respect to the number of telephones served, ranging all the way from New York City with its more than 3,000,000 telephones and 300 central offices to small villages and rural communities with perhaps a few score or a few hundred telephones.

In the 1920's when the Bell System embarked upon its program of converting local offices to dial operation, each exchange or city was in general an entity unto itself. Customers dialed local calls within their own city but all calls involving a toll or multiunit charge required handling by operators for timing and ticketing. There was no advantage, therefore, in making a numbering plan for a given city more comprehensive than required to serve the telephones and central offices in that city with a suitable allowance for the expected growth. Thus there were formed a multitude of local dial communities, large and small, within which customers could dial their own calls and connections between these telephone communities were established by operators.

Over the years these basic numbering plans which were established originally for local dialing have in many of the cities proved inadequate to furnish as many office codes as later events have shown are required. This is due to a variety of causes. The station growth in many places has outstripped all expectations and the number of central offices required to serve this unprecedented demand for service consume many more office codes than the original plans provided for.

In many places local service areas were changed so that customers could call into contiguous exchanges at local rates. To enable customers to dial into these nearby places the original numbering plans required expansion to include this increased number of offices. In addition, with the advance in the telephone art many cities introduced equipment for automatic charging on multiunit and short-haul toll calls so that customers could dial such calls directly instead of placing them with an operator for completion. In order to enable customers to dial these calls, it was necessary to expand the original city numbering plans to encompass wider and wider geographical areas.

In expanding the various types of numbering plans to serve a larger number of central offices than were originally anticipated, various expedients were resorted to. In the largest cities having 3-letter office codes a numeral was substituted for the third letter thus very materially increasing the code capacity from about 325 to about 500 and making it possible to form a number of codes using the same office name. The name CANal for example, instead of serving but one office may serve a number of offices, CAnal 2, CAnal 3, CAnal 4, and so forth. In the medium sized cities having 2-letter codes, expansion meant adding a digit to the code to all or in some cases to only a part of the offices in the city.

The 5-digit places usually were expanded by adding a digit to some of the numbers so that some of the telephones had five digits and others six digits in their numbers.

As a result of choosing originally a numbering plan which at the time seemed adequate and most suitable for the city involved and in many cases being forced to expand to meet changing needs, the Bell System now has a considerable variety of different numbering plans. These numbering plans, illustrated in Table I, all are adequate to serve the present local dialing needs for the cities in which they appear.

Having reviewed the numbering plan situation as it exists today in the various cities and towns, let us turn to the problem of handling toll calls. Under ringdown operation there is an operator at the outward toll center where the call originates and another operator at the terminating or inward toll center. On built-up toll connections there are additional operators at each intermediate toll switching point. The inward toll operators, who are familiar with the numbering plans in the offices served by their particular toll center, can be relied upon to connect to the desired station even though there is uncertainty on the part of the calling customer or the outward toll operator regarding the precise pronunciation or spelling of the name of the called office or the particular form of numbering system used at the called city.

Under operator toll dialing the inward operator is replaced by dial switching equipment under the control of the outward operator; hence the outward operator has no one to rely upon but herself in completing a toll connection to a distant city. With the present method the operator dials a code for each circuit group in the connection followed by the number of the called party which may consist of any number of digits from three to seven. The operator must refer to her position bulletin or to a routing operator for the correct circuit group codes unless she happens to remember them. Where the office to be reached has central office names, the operator must rely on routing information to determine how many letters of the name are to be dialed. The great variation in the number of digits to be dialed on different calls is a source of some difficulty and confusion to the operators.

Paper 52-203, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted May 23, 1952; made available for printing May 23, 1952.

W. H. Nunn is with the American Telephone and Telegraph Company, New York, N. Y.

Table I. Numbering Plans in the Bell System

Place	Directory Listing	Customer Dials	Ordinarily Referred to as
Philadelphia, Pa.	LOcust 4-5678	LO 4-5678	Two-five
Los Angeles, Calif.	PArkway 2345 and REpublic 2-3456	PA 2345 and RE 2-3456	Combined two-four and two-five
Indianapolis, Ind.	MArket 6789	MA 6789	Two-four
El Paso, Tex.	PRospect 2-3456 and 5-5678	PR 2-3456 and 5-5678	Combined two-five and five digit
San Diego, Calif.	Franklin 9-2345 Franklin 6789	F 9-2345 F 6789	One letter, four and five digit
Des Moines, Iowa	4-1234 and 62-2345	4-1234 and 62-2345	Combined five and six digit
Binghamton, N. Y.	2-5678	2-5678	Five digit
Manchester, Conn.	5678 and 2-2345	5678 and 2-2345	Combined four and five digit
Winchester, Va.	3456	3456	Four digit
Ayer, Mass.	629 and 2345	629 and 2345	Combined three and four digit
Jamesport, N. Y.	325	325	Three digit

The present system of operator toll dialing by which operators use codes depending upon the routes to reach a desired destination, is a great improvement over the old manual handling methods. However, with the introduction of more modern toll switching facilities and the nationwide extension of toll dialing, it was realized that an improvement in the methods for dialing toll calls to distant cities was essential in order to realize the maximum speed and accuracy inherent in toll dialing.

These handicaps in the present toll dialing methods are to be overcome by establishing a nationwide numbering plan covering the United States and Canada by which each of the more than 20,000 central offices in the two countries is to be given a distinctive designation which identifies that particular office and that office only. This designation is to consist of two elements, a regional or area code and a central office code. Any outward toll operator, wherever located, will use that same designation in reaching that office through the dial toll switching network.

In a sense, all of the thousands of offices involved are to be treated as though they were contained in one huge multioffice city. Toll operators will use the area code and the office code in reaching an office situated outside her own numbering plan area, while on calls to points within her own numbering plan area she will dial only the number as listed for toll in the directory. In principle the method employed is to divide the two countries geographically into numbering plan areas and to give each of these areas a distinctive code, see Figure 1. Within each of these numbering plan areas each office will have a code unlike that of any other office in the same numbering plan area and also unlike any area code. Hence for toll-dialing purposes each office will have an area code and central office code which will form a combination unlike that of any other central office in the two countries.

In this geographical division into numbering plan areas, border lines between states and between Canadian provinces generally have been used as numbering area boundaries. Since about 500 central offices are the maximum number which can be served in a numbering plan area, it is necessary to divide the larger and more populous states and provinces into two or more areas making, of course, due allowance for growth. New York state with the largest number of central offices is divided into six numbering plan areas; Pennsylvania, Illinois, Texas, and California have four areas each. Other divided states have three or two areas depending upon the number of offices to be served. Approximately 90 areas are being provided, with 14 states and two provinces served by two or more numbering plan areas, the remaining states and provinces by one area each.

In fixing the intrastate numbering plan area boundaries of subdivided states, among other considerations effort was made to avoid cutting across heavy toll traffic routes in order to have as much of the toll traffic as possible terminating in the area in which it originated. The advantage of arranging the numbering plan areas in this manner is readily apparent since on this traffic which does not pass an area boundary the area code is not required.

Let us now consider the composition of the area codes. As indicated previously they must be of a type which will enable the switching equipment to distinguish them from the codes of central offices.

On the telephone dial plate letters are assigned only to the dial positions 2 to 9, inclusive (on some dial plates a Z appears on the 0 position but the Z is never used in a central office code), hence any office code will always avoid a 1 or a 0 in the first two places. The digits 1 and 0 can, therefore, be used in area codes to distinguish these from office codes. It is not practical to use them as initial digits of area codes since customers dial 0 to reach operators and the local dial equipment is arranged to ignore an initial 1 for technical reasons. A 1 or 0 in the second place, however, can be employed in an area code without conflicting with any central office codes or interfering with any existing practices. Accordingly, the area codes will consist of three digits with either a 1 or a 0 as the middle digit, 516, 201, and so forth. A few codes of this type are now in use, leaving a practical total of 152 of these area codes available as compared to approximately 90 assigned to our present numbering plan areas. This will provide a comfortable spare for additional future numbering plan areas or possibly for reaching overseas points which may be incorporated later into the toll dialing network.

As shown on the chart, states and provinces such as Montana or Alberta which are contained in a single numbering plan area will have area codes with a 0 as the middle digit to distinguish them from areas in divided states such as Texas where the middle digit will be a 1. This is to enable toll operators to differentiate between the two classifications of areas. On calls to single area states the operators will always know that every call to the state in question uses the one area code, whereas on calls to subdivided states additional information will be required to determine which of the several area codes should be employed to reach the particular destination. It is proposed to show on the operator position bulletin the codes of all single area states and the codes of all frequently called cities in multiarea states. The area codes of the less frequently called places in the multiarea will be obtained from a routing operator.

Within each numbering plan area each of the 500 or fewer offices are to be given a 3-digit office code which will be different from that of any other office code in that same area. Ultimately each central office will be given a 2–5 type of number consisting of an office name and five numerical digits, such as LOcust 4-5678, illustrated for Philadelphia. In the larger cities customers will dial seven digits, LO 4-5678, on local calls to numbers in the same exchange. In many of the smaller places the customers on local calls will dial only the numerical digits, the office name being employed for toll dialing purposes only.

Considering the thousands of central

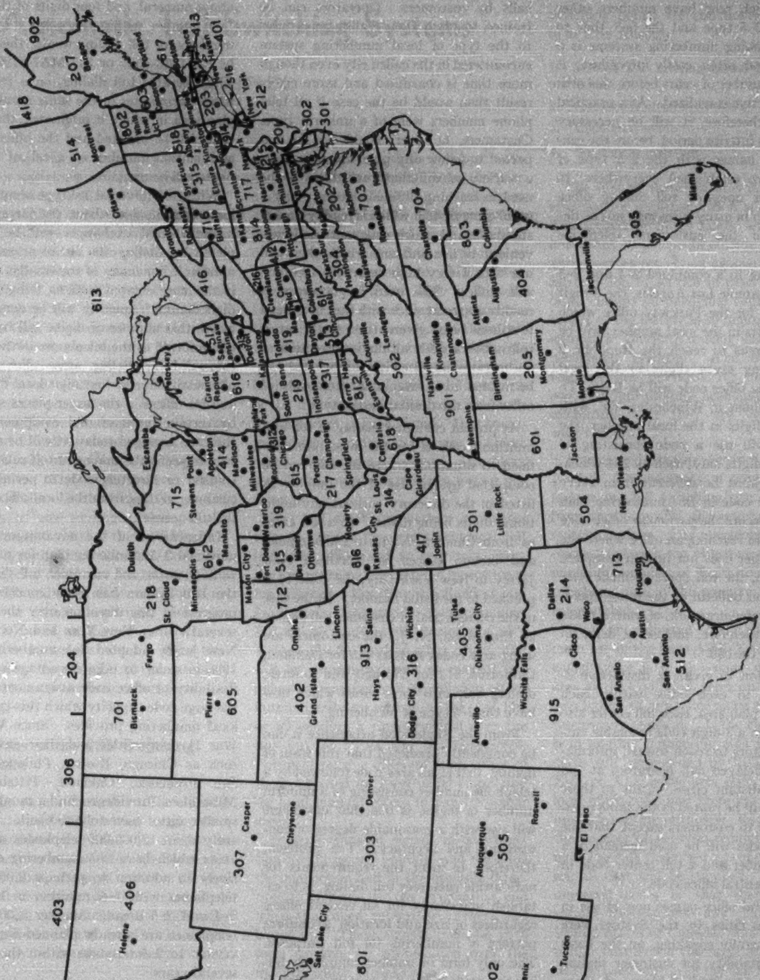

902
207
Bangor
Portland
617
Boston
413
401
418
603
802
White River Jct.
Concord
203
516
New Haven
New York
212
514
Montreal
518
Albany
315
Syracuse
914
Kingston
607
Elmira
201
Newark
302
301
609
Baltimore
717
Scranton
Harrisburg
215
Philadelphia
Ottawa
716
Rochester
Buffalo
814
Kane
Altoona
412
Pittsburgh
202
Washington
703
Richmond
Norfolk
Roanoke
304
Clarksburg
Charleston
Huntington
704
Charlotte
803
Columbia
Augusta
404
Atlanta
305
Jacksonville
Miami
416
Toronto
613
216
Cleveland
Canton
614
Columbus
513
Dayton
Cincinnati
419
Toledo
517
Saginaw
Lansing
313
Detroit
616
Grand Rapids
Kalamazoo
Petoskey
502
Louisville
Lexington
Knoxville
Chattanooga
615
Nashville
901
Memphis
205
Birmingham
Montgomery
Mobile
601
Jackson
504
New Orleans
219
South Bend
317
Indianapolis
812
Terre Haute
Evansville
312
Chicago
Highland Park
414
Milwaukee
Madison
Appleton
Rockford
815
Peoria
217
Champaign
Springfield
618
Centralia
Cape Girardeau
314
St. Louis
Escanaba
715
Stevens Point
319
Waterloo
515
Fort Dodge
Des Moines
Ottumwa
Mason City
Moberly
816
Kansas City
417
Joplin
501
Little Rock
713
Houston
214
Dallas
Waco
512
Austin
San Antonio
San Angelo
915
Cisco
Wichita Falls
El Paso
405
Tulsa
Oklahoma City
316
Wichita
Dodge City
913
Salina
Hays
402
Omaha
Lincoln
Grand Island
712
612
Minneapolis
Mankato
Duluth
218
St. Cloud
204
Fargo
701
Bismarck
605
Pierre
306
403
406
Helena
307
Casper
Cheyenne
303
Denver
505
Albuquerque
Roswell
801
Salt Lake City
602
Phoenix
Tucson
208
Boise
604
206
Spokane
Seattle
503
Portland
702
Reno
714
San Diego
Los Angeles
213
916
Redding
Sacramento
415
San Francisco-Oakland

Figure 1. Numbering plan areas with codes

offices which now have numbers other than the 2–5 type and the fact that to change existing numbering systems is a difficult and often costly procedure, it will be a number of years before this ultimate objective is realized. As a practical measure, therefore, it will be necessary during this interim period, before the central office names with the 2–5 type of number are established everywhere, to employ for operator toll dialing office codes which in many cases may not be derived from the customers' telephone number.

In dialing to a combined 2–4 and 2–5 city, for example Los Angeles, the 3-digit office code for the Parkway office which has six digits in the local number, will be PAR, whereas to reach the Republic 2 office having seven digits in the local number, the office code will be RE2. To call a telephone in Winchester, Va., with only four digits in the local number, the operator will use a code consisting of numerical digits only, such as 294 which, of course, must be different from every other office code in this numbering plan area. To secure the particular office code to be used in reaching an office where the called number does not furnish complete information, the toll operator must refer to a position bulletin or the route operator. This reference work, of course, takes time and, therefore, imposes a delay in completing the call.

In addition to giving a distinctive 3-digit code to each office within each numbering plan area, each toll center also will be given a 3-digit code to enable outward operators to reach inward information, and delayed call operators at toll centers in distant cities. Calls to these operators will be routed in the same manner as calls to customers except that the operator codes will be used instead of a station number and a toll center code in place of a central office code.

The central office names now in use in the various cities in the System were chosen, generally speaking, on the basis of their suitability for customer dialing within the city itself. Many of these names are unfamiliar words to operators and customers in distant cities and the use of these names contributes materially to the operator dialing errors. This situation is gradually being corrected by using for new offices names from a System-approved list and replacing existing names which experience has shown to be particularly troublesome by names from this list.

While numbering plans are important in operator toll dialing, they play an even more essential part in the dialing of toll calls by customers. Operators can be trained to adapt their dialing procedures to the type of local numbering system encountered in the called city even though more time is consumed and more errors result than would be the case if all telephone numbers were of a uniform type. Customers, however, could not be expected to follow any plan which requires a variety of different procedures to be used in reaching different cities. Only a numbering system which is readily understandable and which customers find convenient to use and one which they can use with a very high degree of accuracy will suffice. The need for accuracy is readily apparent since with the customer's telephone being given access to the intertoll network without the intervention of an operator, a call which is misdialed can be routed to a telephone thousands of miles from the desired destination.

At present customer dialing of toll and multiunit calls is for the most part confined to situations where the call can be completed by the use of the number as listed in the directory without any additional digits being dialed. In a few cases as from Camden, N. J., to Philadelphia, and certain offices in Northern New Jersey to New York City, the code 11 is prefixed to the listed number. In the case of the current trial of customer toll dialing at Englewood, N. J., the customers are using area codes such as 415 for Oakland, California, 312 for Chicago, and so forth, dialing only into those cities which now have the 2–5 type of numbering.

From the Englewood experience it can be confidently predicted that this form of dialing, that is, an area code followed by a telephone number consisting of a uniform number of digits, is one that customers will use with a reasonable degree of convenience and accuracy. The problem, therefore, to meet the requirements for nationwide customer toll dialing, is to establish universally for all central offices regardless of size and location a uniform pattern of numbering for toll purposes. The only form of number completely filling the needs is the 2–5 system, which is that used in the largest cities today.

Accordingly, in order to implement the program for customer dialing of toll calls on a nationwide basis, it will be necessary to place all telephone numbers on a 2–5 basis with the code of each office different from that of every other office in the same numbering plan area. Thus each of the 50,000,000 telephones in the United States and Canada will have, for toll dialing purposes, a distinct identity consisting of ten digits; a 3-digit area code, an office code of two letters of an office name and a numeral, and four digits of the station number within the office. Typical numbers for toll dialing would, therefore, be 601-CA3-4567 or 317-MA7-6789. As with operator toll dialing, on a toll call which terminates in the same numbering plan area in which it originates, the area code will be omitted and the office code and station number—a total of seven digits will be used.

With this universal 2–5 type of number, local calls in and about the larger and medium sized exchanges will be completed by dialing the entire seven-digit number. For many of the smaller places in the more isolated sections, 5-digit or 4-digit dialing frequently will be employed where this number of digits will be adequate for all of the telephones in the customers' local dialing area. For these offices with five or four-digit local dialing and for offices in the larger places served by certain types of dial equipment, as they are arranged today, it will be necessary to prefix the dialing of toll calls by a transfer or directing code to permit the customer getting from the local office into the toll network.

Independent of the advantages of a universal 2–5 numbering plan for nationwide operator and customer toll dialing, the Bell System has made considerable progress in this direction over the past several years. New York and Northern New Jersey adopted 2–5 numbering in 1930 in order to take advantage of the flexibility of office code assignments and the large code capacity which this type of local numbering provides. Since World War II many cities and their environs such as Chicago, Boston, Philadelphia, San Francisco, Oakland, Pittsburgh, Milwaukee, Providence, and a number of smaller cities have followed suit. Presently about 12,000,000 telephones are in areas which have 2–5 numbering exclusively in addition to perhaps 2,000,000 telephones with 2–5 numbers in mixed 2–4 and 2–5 areas. Another 5,000,000 telephones are already planned for conversion to 2–5 numbers within the next several years.

The entire program will take many years to realize but it is one which must be accomplished in order to achieve the best results in operator toll dialing and make it possible for a customer at any telephone in the United States and Canada to reach a telephone anywhere in the two countries by dialing without the assistance of an operator.

No Discussion

Automatic Toll Switching Systems

F. F. SHIPLEY
MEMBER AIEE

THIS paper deals primarily with the major switching centers required for the nationwide automatic switching plan. These are called control switching points (CSP's) and are supplied with switching equipment endowed with great versatility and a high order of mechanical intelligence. Mr. Pilliod's paper[1] explains how for purposes of circuit layout and routing, they are assigned different rankings as follows, starting with the lowest ranking: primary outlets (PO's), sectional centers (SC's), regional centers (RC's), and one national center (NC). Substantially the same equipment is to be provided for all of these centers so that they all will have inherently the same capabilities. They will, however, differ greatly in size. In the United States and Canada, as now envisaged, there will be somewhat under 100 of these CSP's.

The system which the Bell Telephone Laboratories developed for use at CSP's and which embodies all of the features required at those important switching points is based on the toll crossbar system[2] now in service and has been constructed by the addition of the necessary CSP features to the basic structure of that system.

Functions of the CSP Switching System

The system is designed to be suitable for location in either a step-by-step or a panel-crossbar local area. In addition to the functions required for operation as a CSP, it must, of course, perform the normal toll switching functions required of any system for switching the toll traffic characteristic of the locality it serves. These may be stated very briefly.

ORDINARY TOLL SWITCHING FUNCTIONS

It accepts calls either directly from operators or from senders in distant offices. In the interest of economy, it accommodates itself to the signaling language the operator's position or sender is equipped to deliver. Calls from operators may be either in the dial pulse (DP) or multifrequency[3] (MF) form. Calls from senders will be in the MF form. DP pulsing is the decimal type delivered directly by the dial and is at the rate of about one digit per second. MF pulsing represents a particular digit by a combination of two out of five frequencies in the voice range; it uses one of these frequencies in combination with a sixth frequency to produce a signal indicating the beginning of pulsing, and a different one of the five in combination with the sixth for an end of pulsing signal. It is transmitted from senders at the rate of about seven digits per second. Operators usually key at the rate of up to two per second.

The toll switching system completes calls to various types of mechanical toll and local offices and to operators, using the form of signaling dictated by economy for each call. For distant toll offices and local offices using step-by-step equipment DP will be transmitted; for other CSP's and usually for local crossbar offices, MF will be transmitted and at manual toll offices an operator will be called in either automatically on seizure of the toll line or by sending a ringing signal over the line but no pulses will be transmitted. Forms of pulsing different from either of these are used for local panel offices and for local manual offices in panel-crossbar areas.

It must transmit signals in one direction for initiating, holding, and releasing the connection and in the opposite direction to indicate to the originating end when the called subscriber answers and hangs up. These signals must be in a form suitable for propogation over the medium which carries them.

It must exercise control over the amount of amplification of voice currents introduced at the switching point so that a proper grade of transmission will be furnished.

All of these functions are performed by toll crossbar systems already in service. The features that distinguish the new system are those peculiarly characteristic of CSP operation.

CSP FUNCTIONS

The following features which will be built into the equipment at control switching points are commonly referred to as CSP features:

1. Storing and sending forward digits as needed.
2. Automatic alternate routing.
3. Code conversion.
4. Six-digit translation.

The first of these features is basically essential for implementation of the plan. The second produces faster service and important economies in outside plant. It also provides protection against complete interruption of service in case of failure of all circuits on particular routes. This aspect of the feature is so important that automatic alternate routing also may be considered essential. The other two features are provided for reasons of economy and produce economies of such magnitude that they are very much worth while.

Storing and Sending Forward Digits as Needed. The necessity of providing this feature in CSP switching systems arises from the nature of the numbering and switching plans. The numbering plan[4] is constructed with the objective of using a minimum number of digits to give each telephone user in the country a distinctive number.

Numbers delivered to the CSP equipment are in the form ABX-XXXX if the called place is in the same numbering area as the CSP. AB represents the first two letters of any office name and X represents any numeral. If the called place is in another numbering area this set of digits will be preceded by X0X or X1X. X0X or X1X is the area code, ABX the local office code, and these are the digits used for routing purposes. Regardless of the number of switches required to complete the call, these two sets of code digits are all that will be supplied. They are universal codes in that they identify specific destinations—any place in the United States or Canada—and for a particular destination the same set of digits will be used wherever the call may originate. All CSP's must, therefore, be able to advance a call toward the same place when the same set of digits is received.

To make use of destination codes possible, each CSP must store the digits as received and pass along to the next point whatever digits may be required there for advancing the call. If the next point is a CSP not in the home numbering area of the called place, the complete 10-digit number will be sent forward. If it is a CSP in the home numbering area of the called point, the area code will be dropped and the remaining seven digits will be sent

---

Paper 52-204, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted May 23, 1952; made available for printing May 23, 1952.

F. F. SHIPLEY is with the Bell Telephone Laboratories, Inc., New York, N. Y.

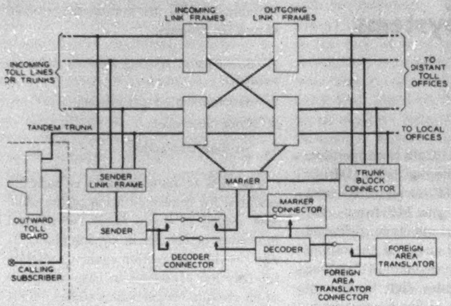


Figure 1. Schematic diagram of crossbar switching system for CSP's

forward. That CSP may in turn complete to a local office directly, dropping the office code, or through a step-by-step TO (tandem outlet) or TC (ordinary toll center), substituting arbitrary digits for the area or office code, thereby exercising the third of the listed CSP features.

Automatic Alternate Routing. The system is arranged to offer the maximum number of alternate routes possible under the switching plan. As explained by Mr. Pilliod[1] a maximum of five alternates actually will be used. This number is possible, of course, only at PO's since higher ranking CSP's have fewer CSP's above them in the final chain.

Code Conversion. This refers to the ability to substitute one, two, or three arbitrary digits for the area code, the office code, or both. It is economically important to be able to do this because it makes it possible to work with the step-by-step equipment extensively used in local offices and in toll offices in TC's or TO's without the changes in local numbering plans or rearrangements—and in some cases extra selectors—required for the step-by-step TC's or TO's to use ABX codes for routing purposes. Even though eventually all customers are listed as ABX-XXXX and TC's are arranged to use the listed number for routing the calls, this will not be accomplished for some time. Moreover, after such arrangements are in effect there will still be need for code conversion, particularly for routing calls through TO's. Many combinations of digit dropping and substitution are required to cover all possible cases.

Six-Digit Translation. When a CSP receives a 10-digit number, it is sometimes sufficient to translate only the area code digits and sometimes necessary to translate both the area and office codes. If all points in the called area are reached by the same route out of the particular CSP concerned, the area code will suffice for selection of the route. If some points are reached by one route and others by one or more different routes, the office code also must be translated to determine which route should be selected.

Basic Arrangement

In the CSP switching equipment talking connections are established through crossbar switches.[5] Incoming and outgoing toll lines and toll connecting trunks are terminated on crossbar switch frames with linkage between them to provide full access. The switches are controlled by equipment common to the office, each item of which is held only long enough to perform its task in setting up the connection.

The major items of common control equipment are senders, markers, decoders, and translators. The basic functions of the senders are the same as in other common control systems, that is, registering incoming digits and sending them out as directed. A departure from prior practice is made in the design of the marker. In other crossbar systems the marker is the principal seat of the mechanical brains. It not only controls the actual establishment of the connection but also does the translating to determine what connection should be established and what information should be passed to the sender for further disposition of the call. In this system, the marker still controls the actual setting up of the connection, but it acts on instructions received from the decoder where the major portion of built-in intelligence resides.

The decoder accepts code digits from the sender, translates them, makes selection of alternate routes, and gives instructions to markers and senders to enable them to carry out their assignments. To do the translating job, the decoder has one, and in some cases two, translators permanently associated with it and in addition has access to a common group of translators called foreign area translators which can be used by all decoders as required.

The relationship of the principal elements of the system to each other is depicted in the schematic diagram, Figure 1.

Method of Operation

The manner in which the various elements of the CSP system and the CSP systems at various locations co-operate to implement the nationwide switching plan may be understood best by following the progress of a call which demands the exercise of the characteristic CSP functions.

Assume an outward operator in Atlanta has received a station-to-station call for a subscriber in Monticello, Maine, whose number is ACademy 4-2345, that Monticello is a tributary of Houlton, a step-by-step TC, that Bangor is a step-by-step TO serving Houlton as its home TO, and that the circuit groups provided are as indicated in Figure 2. The dotted lines represent high usage groups and the solid lines final groups.

The Atlanta operator plugs into a tandem trunk to the toll crossbar system in Atlanta, thereby causing a sender to be attached to the trunk through the sender link frame. This causes a lamp signal to be displayed to the operator, indicating that she may key the number. She keys 207-AC4-2345 plus a start signal (signifying end of keying) into the sender and leaves the connection to handle other calls. She will give this call no further attention until the lamp associated with the cord circuit used in establishing it signifies either by flashing that the call was not completed due to a busy condition of the called line or to circuit congestion, or by going dark that the called subscriber has answered and she should start timing the call.

As soon as the area code 207 is received by the sender, it calls for a decoder and gives it the code. The decoder, by means of a self-contained translator, finds that the area code is sufficient for routing purposes, that the first choice route is by way of Boston, the second New York, and the final St. Louis. Without consulting other circuits it will know in which of these groups an idle circuit may be found. Let us assume that the circuits to Boston are all busy but that there are one or more idle circuits in the New York group. The decoder calls for a marker and tells it which group of leads to test, and also causes the sender to be connected to the particular marker it has selected.

The marker, following instructions from the decoder, is connected to the appropriate trunk block connector. This is one of a group of common circuits giving access to "blocks" of trunks for allowing the marker to locate an idle trunk. The marker examines the test leads of the individual toll lines to New York, and as soon as it has selected an idle circuit it so informs the decoder. The decoder then tells the sender to send all digits forward by MF and leaves the connection to accept another call. This information is relayed from the decoder to the sender by way of the marker. The work time of the decoder has been in the order of 1/2 second.

The marker determines the identity of the frames on which the incoming and outgoing circuits are located, finds an idle path between the two circuits, and sets up the connection. After checking the path through the switches to be sure that there are no troubles, it notifies the sender that its task has been completed and then leaves the connection. Its work time has also been in the order of 1/2 second.

In the meantime other digits have been coming in to the sender but it does not wait for all of them to arrive before advancing the call. When the marker selected the circuit to New York a signal was sent forward immediately to summon a sender in the New York switching system. The process of attaching the sender in New York was carried on concurrently with the establishment of the connection through the switches in Atlanta.

When the New York sender is attached, a signal is sent to the Atlanta sender to advise it that pulsing may proceed. It immediately sends the area code 207 to New York by MF pulsing and follows it with the remaining digits of the called number, AC4-2345, as they are received from the operator, ending with a start pulse, and then leaves the connection. All common control equipment in Atlanta is now free.

In New York, as soon as the Maine area code is received, it is submitted to the decoder. Upon examination of the code, the decoder finds that it is insufficient for routing purposes. New York has a direct circuit group to Portland over which traffic to some offices in Maine is routed, but other offices are reached through Bangor by way of Boston. In order to determine which route to take the decoder must know what office is desired. Therefore, it gives the sender a signal saying "come again when you have six digits" and leaves the connection. When the sixth digit arrives the sender again calls for a decoder and gives it the complete code 207-AC4.

The decoder again translates the area code, which now directs it to the foreign area translator which serves the Maine area, and submits the complete code to that translator. From the ensuing translation it learns that the route is by way of Boston and that all digits should be sent forward by MF. It then calls for a marker and releases the foreign area translator.

Subsequent operation is the same as previously described for Atlanta and the complete 10-digit number now arrives at Boston. At that point both codes are again translated since Boston also has a choice of routes to Maine, and the route to Bangor is selected. The translating equipment in Boston knows that Bangor is in the Maine area and that the area code will, therefore, not be needed. However, since Bangor is a TO having no senders, the Boston sender must pulse forward all of the digits needed to complete the call through switches in Bangor, Houlton, and Monticello. It is assumed that Houlton is arranged to route the call to Monticello on receipt of the digits AC4. Numerical digits 2345 will route the call through the Monticello switches to the called customer's line. These digits are all registered in the Boston sender but the digits required to switch the call through Bangor are not and must be supplied. An arbitrary set of digits beginning with 1 can be used for this purpose since no office code begins with 1 and there will, therefore, be no conflict.

The decoder in Boston, therefore, gives the sender the proper set of arbitrary digits, say 16, to be placed ahead of the office code AC4. The sender sends forward by the DP method 16-AC4-2345 driving switches in Bangor, Houlton, and Monticello to the called subscriber's line, and ringing starts automatically. The talking connection is now established and the common control equipment at all intermediate points is free.

When the called subscriber answers, the Atlanta operator's cord lamp is extinguished. When he hangs up, the lamp lights to denote end of conversation. The removal of the operator's cord automat-

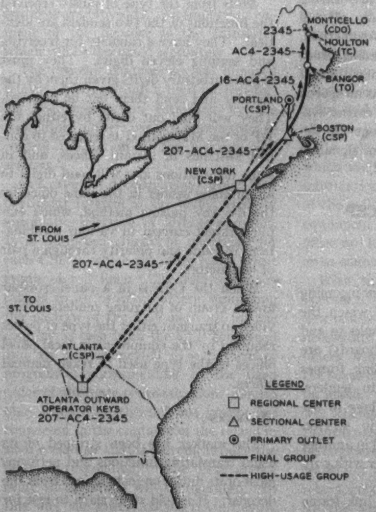
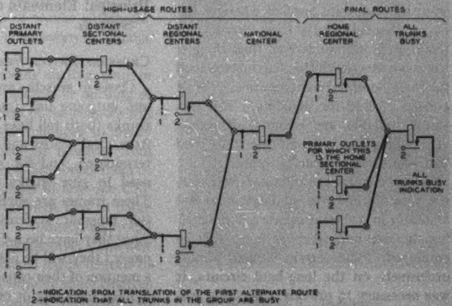

2345
MONTICELLO (CDO)
HOULTON (TC)
AC4-2345
16-AC4-2345
BANGOR (TO)
PORTLAND (CSP)
BOSTON (CSP)
207-AC4-2345
NEW YORK (CSP)
FROM ST. LOUIS
207-AC4-2345
TO ST. LOUIS
ATLANTA (CSP)
ATLANTA OUTWARD OPERATOR KEYS 207-AC4-2345
LEGEND
REGIONAL CENTER
SECTIONAL CENTER
PRIMARY OUTLET
FINAL GROUP
HIGH-USAGE GROUP

Figure 2 (left). Call from Atlanta, Ga. to Monticello, Maine

Figure 3 (below). Alternate route array for the decoder at a sectional center


HIGH-USAGE ROUTES
FINAL ROUTES
DISTANT PRIMARY OUTLETS
DISTANT SECTIONAL CENTERS
DISTANT REGIONAL CENTERS
NATIONAL CENTER
HOME REGIONAL CENTER
ALL TRUNKS BUSY
PRIMARY OUTLETS FOR WHICH THIS IS THE HOME SECTIONAL CENTER
ALL TRUNKS BUSY INDICATION
1
2
1-INDICATION FROM TRANSLATION OF THE FIRST ALTERNATE ROUTE
2-INDICATION THAT ALL TRUNKS IN THE GROUP ARE BUSY

ically releases the entire connection, the release of each link causing the next in line to release.

In setting up this call all of the characteristic CSP features were employed, automatic alternate routing in Atlanta, 6-digit translation in New York and Boston, digit storing, and variable spilling at all CSP's with substitution of arbitrary digits for the area code at Boston.

Transmission

All talking connections through the CSP system are made on a 4-wire basis, that is, separate pairs of conductors are provided for transmission in the two directions. This is done in order to simplify the problem of maintaining satisfactory balance so that the loss introduced by extra links in a connection can be held to a minimum value. The importance of this feature is emphasized by the fact that the switching plan permits as many as eight intertoll trunks to be connected in tandem for the completion of a call.

The advantages of 4-wire switching were fully explained in a paper[2] on the toll crossbar system now in service.

Signaling

In following the progress of the call from Atlanta to Monticello, Maine, it was observed that besides the transmission of information in the form of digits, it was necessary to pass a number of control and supervisory signals over the toll lines. These included seizure and disconnect signals in the forward direction and switchhook supervisory signals and sender attached signals in the reverse direction. On some calls it is necessary also to send flashing signals to indicate busy lines or trunks and ringing signals in either direction when operators are called in at intermediate or terminating points to assist in establishing connections.

For the early toll dialing installations, the signaling method most widely used was the composite method whereby signaling channels for the three circuits of a phantom group are derived from three of the conductors with the fourth being used for earth potential compensation. Direct current is used for signaling. This is a simple, reliable, and economical method of signaling and will continue to be used on circuits where it can be applied.

Where circuits are obtained from carrier systems, however, conductors are not available in sufficient numbers for signaling channels and other methods must be employed. Since carrier is used almost exclusively on the long-haul circuits, it was necessary to provide a signaling system to accompany it before toll dialing could be expanded beyond networks of limited range. To meet this situation a system[6] using a frequency of 1,600 cycles was developed and has been in service since 1948. Signaling is done by application and removal of the 1,600-cycle signaling current. The system is used in the same manner as the composite signaling system, to carry dial pulses as well as supervisory signals when used on circuits that require it. The set of leads brought out of the signaling unit are identical in function to those brought out of the composite signaling unit so that toll-line relay circuits will operate in the same manner with either type of signaling.

Since 1,600 cycles is in the voice range, the signaling current can be carried over the same channel that carries the speech current; but the signaling circuits must, of course, be protected against false operation due to speech, and precautions must likewise be taken to insure that the signaling tone does not interfere with speech. Protection against interference between signaling and speech is more difficult at 1,600 cycles than at higher frequencies because there is more energy in voice currents at the lower range. That value was chosen, nevertheless, so that it would be possible to operate over the narrow-band circuits that were established to relieve shortages occasioned by the war.

A new 2,600-cycle system to be used only on the broader band circuits has since been developed. It is simpler and more economical than the 1,600-cycle system. The older carrier systems, having been designed when practically all toll operation was by the ringdown method, made no provision for signaling since that was all done by short applications of the 1,000 cycles when there was no speech on the line. Some of the new carrier systems for short-haul applications are designed to provide their own signaling channel for each voice channel.

Principal Elements of the CSP System

Crossbar Switch Frames

Crossbar switches are used for incoming and outgoing link frames on which the trunks (both toll lines and trunks to and from local offices and switchboards) are terminated, and for sender link frames used to give trunks access to senders. These frames are similar to those in the toll crossbar systems now in service. Since they have been described in another paper,[1] they will be passed over with only a mention of their capacity.

Each incoming or outgoing link frame normally has terminals for 300 trunks. As many frames are provided as required for the size of the office. In the smaller offices, one train of switches with complete interconnection of incoming and outgoing frames is provided. In the larger offices, two trains each with its own set of markers are provided. When this is done, the incoming trunks are multipled to both trains and an extra build out bay is provided on the incoming frame to provide 400 terminals per frame. Since each train has a theoretical limit of 40 incoming and 40 outgoing frames, the maximum size of an office is theoretically 80 of each. Practical considerations, however, such as the number of markers that can be operated efficiently in a group and the maximum size office it is feasible to operate as a single administrative unit will limit an installation to about 60 incoming and 60 outgoing frames.

The sender link frame gives 100 trunks access to 40 senders.

Senders

Two separate groups of incoming senders are provided, one to receive DP and the other MF pulsing. Whether the system is installed in a step-by-step or a panel-crossbar area, both groups of senders always will be needed. MF will be received from senders in other CSP's and from switchboard positions. DP will be received from switchboard positions at TC's not equipped to send MF and in some cases from dialing A boards in the local area of the CSP itself.

Aside from the type of pulses received the functions of the two senders are identical. They have a capacity for receiving and sending eleven digits. They must register arbitrary digits given them by the decoder and send them out as directed. They will send out digits by either the DP or the MF method as required to control switches in distant offices, and in some installations also will send digits to an outgoing sender in the same office by the d-c key pulsing method, which employs direct current in various combinations of value and polarity through a pair of conductors.

When the CSP is in a panel-crossbar area a group of outgoing senders is provided to transmit either the type of pulses required by the equipment in local panel offices or the type used to reach manual offices.

Markers

The marker has been stripped of its usual translating functions and performs most of its duties on instructions from the decoder. It is told what leads to test for

Figure 4. Card translator

idle circuits and where they are to be found in the trunk block connector, but having found an idle circuit it carries on the process of setting up the connection independently of the decoder. Having contact with both the incoming and outgoing trunks through connecting circuits, it determines what frames they are located on, connects itself to those frames, selects a path through them and sets up the connection.

In a single-train office, one group of markers common to the office is provided. In a 2-train office there is a group of markers associated with each train of switches.

DECODERS

A single group of decoders serves the entire office whether one or two trains of switches are provided. An important element of the decoder is the translator which will be discussed separately.

The decoder contains several hundred relays. A large group is used for registering the information furnished by the translator. Others use this information to control the action of the markers and senders.

One group of decoder relays which is of particular interest is the array used for automatic selection of alternate routes. It is composed chiefly of one relay for each CSP to which the office has a direct group of toll lines. The relays are arranged in an orderly pattern simulating the pattern of the CSP network for the country as seen from the CSP concerned and are interconnected in a pattern of progression corresponding to the fixed order of alternate route selection. Group busy leads from the toll line groups are connected to the contacts of the relays in such a manner that if a group is busy the relay corresponding to the next choice route in the chain will be operated. In this way the lowest choice route having an idle circuit will be selected speedily without testing individual trunks of separate groups. The decoder learns from the translator which relay in the array to operate first and the choice of the best route available follows automatically. The principle will be understood readily by reference to the simplified sketch in Figure 3. Contacts not shown on the relays cause the translator to select the route corresponding to the last relay operated in the chain.

TRANSLATORS

The magnitude of the translating job for nationwide dialing led to the decision to develop a new translator operating on a principle radically different from that employed in other crossbar systems. In previous systems, translation is done by relays. The code digits—never more than three—operate a group of relays which cause a single terminal corresponding to the code to be selected. A cross-connection is made between the code point and a route relay associated with the trunk group to be selected. The route relay has a number of contacts which are cross-connected to supply the information required for proper routing of the call. When changes in routing or equipment location of trunks within the office are made, it is necessary to change cross-connections.

With the nationwide dialing plan in operation, routing changes or opening of new offices in one part of the country will necessitate translator changes in many offices, some of them far removed from the scene of the event that forces them to be made. The changes in any one CSP will, therefore, be frequent and to make them by running cross-connections would be cumbersome and expensive. The new translator uses punched cards instead of relays, making it possible to effect changes by the simple process of removing old cards and inserting new ones in a machine. This can be done in a very short time and not only saves labor but requires less out-of-service time for the equipment. Figure 4 shows the machine.

A metal card about 5 by $10\frac{3}{4}$ inches is provided for each area code and also one for each office code that must be translated in a particular CSP, the cards representing destinations. The capacity of a single machine is about 1,000 cards. The cards are lined up in a box as in a filing drawer, with tabs along the bottom of the card resting on select bars which run the length of the box. One hundred and eighteen holes are punched out in all cards in fixed positions so that in the normal condition 118 tunnels are formed from one end to the other. A light source at one end of the box shines through the tunnels upon phototransistors (Figure 5) at the other end but the phototransistors are disabled until, concurrently with the dropping of a card, voltage is applied to them.

All tabs along the bottom of the card are cut off except those which serve to identify the particular card. When a code is presented to the machine a combination of select bars corresponding to the code is lowered. The card having all tabs cut off except those resting on the lowered bars will drop, but all other cards will remain in place. If nothing further were done, the dropping of the card would cut off all light channels; but on each card

Figure 5. Translator

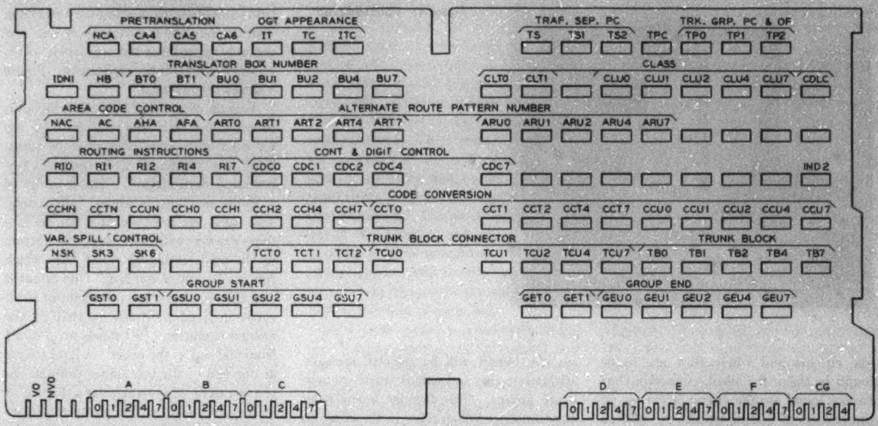

PRETRANSLATION
NCA
CA4
CA5
CA6
OGT APPEARANCE
IT
TC
ITC
TRAF. SEP. PC
TS
TS1
TS2
TPC
TRK. GRP. PC & OF
TP0
TP1
TP2
TRANSLATOR BOX NUMBER
IDN1
HB
BT0
BT1
BU0
BU1
BU2
BU4
BU7
CLASS
CLT0
CLT1
CLU0
CLU1
CLU2
CLU4
CLU7
CDLC
AREA CODE CONTROL
NAC
AC
AHA
AFA
ALTERNATE ROUTE PATTERN NUMBER
ART0
ART1
ART2
ART4
ART7
ARU0
ARU1
ARU2
ARU4
ARU7
ROUTING INSTRUCTIONS
RI0
RI1
RI2
RI4
RI7
CONT & DIGIT CONTROL
CDC0
CDC1
CDC2
CDC4
CDC7
IND2
CODE CONVERSION
CCHN
CCTN
CCUN
CCH0
CCH1
CCH2
CCH4
CCH7
CCT0
CCT1
CCT2
CCT4
CCT7
CCU0
CCU1
CCU2
CCU4
CCU7
VAR. SPILL CONTROL
NSK
SK3
SK6
TRUNK BLOCK CONNECTOR
TCT0
TCT1
TCT2
TCU0
TCU1
TCU2
TCU4
TCU7
TRUNK BLOCK
TB0
TB1
TB2
TB4
TB7
GROUP START
GST0
GST1
GSU0
GSU1
GSU2
GSU4
GSU7
GROUP END
GET0
GET1
GEU0
GEU1
GEU2
GEU4
GEU7
VO
NVO
A
B
C
D
E
F
CG
0 1 2 4 7

Figure 6. Card layout

some holes are enlarged and through these holes the light continues to shine, energizing the corresponding phototransistors. The combination of enlarged holes furnishes all of the information needed for routing the call to the destination represented by the card.

Figure 6 shows the functions of the various groups of tabs and holes. The designations will not appear on the actual cards. Figure 7 shows an actual card prepared for use.

Selecting Tabs—Input Information

The sole use of the information presented to the card translator is to enable it to select the proper card. The information presented is in the form of code digits with accompanying indications of their nature. The information is recognized by cutting off tabs along the bottom of the card in proper combinations.

The groups of tabs labeled A, B, C, D, E, and F are for the 6 code digits. For each digit used two tabs are left since the digits are registered in the sender on a two-out-of-five basis and the leads from the sender will operate the select bars directly. If the card represents an ordinary 3-digit code all tabs will be cut off except two each of the A, B, and C tabs, two of the four CG tabs and perhaps either the VO or NVO tab. The CG (card group) tabs are used in combination to indicate 3-digit, 6-digit, and alternate route card groups. The VO and NVO (via only and not via only) tabs are used when the group of toll lines over which the call will be routed is divided into one subgroup of a transmission grade suitable for only terminal traffic and another subgroup for either terminal or switched traffic. If the card represents an ordinary 6-digit code two tabs will be left in each of the digit positions, and a different pair in the CG group.

Punch Holes—Output Information

The output information from the card translator is recognized in the decoder and marker by relays operated in the combinations set up by enlargement of associated holes in the card. The output from the phototransistors is amplified by other transistors to fire cold cathode tubes which in turn operate the relays.

The pretranslation group on the top line of Figure 7 indicates how many digits the sender must supply for a complete translation. The term "pretranslation" implies that further translation is required. This is not always true. In many cases only the first three digits need to be translated and all information needed for routing the call is supplied by this card. In many cases, the six digits of the area and office code are needed and the routing information will be on another card to be selected later. For certain calls such as calls to operators, only four or five digits are needed. These are treated as 6-digit calls by having the sender supply the extra one or two digits to fill the complement. The NCA hole enlarged means "no come again," that is, three digits are sufficient and translation will proceed. The other holes enlarged mean respectively "come again when you have four, five, or six digits," and no further translation is done until the sender comes back a second time, probably to a different decoder, with an indication that it has the required number of digits.

The OGT appearance holes are used in a 2-train office to tell which train the outgoing trunk appears on and enable the decoder to select a marker in the proper group.

The remaining holes on the top line are for controlling operation of traffic meters.

The translator box number holes in the second line are punched on the area code cards to indicate which machine con-

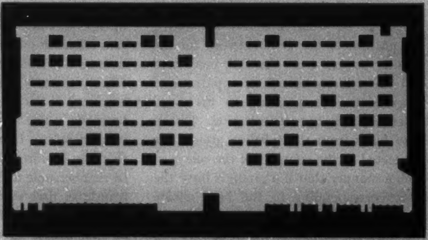


Figure 7. Punched card

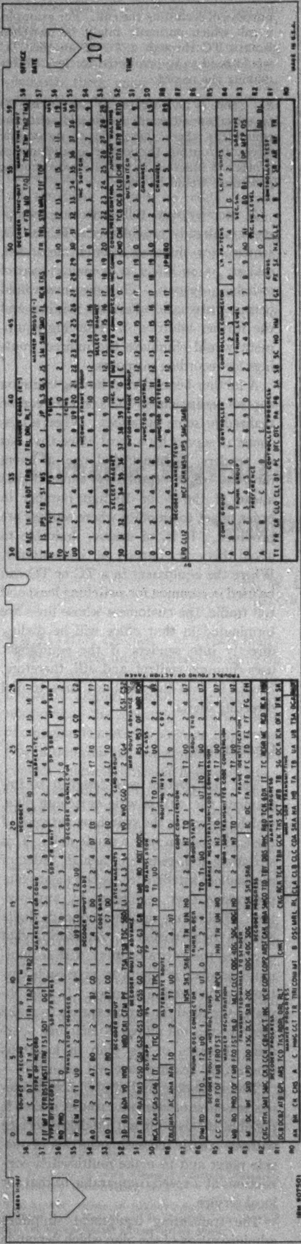


Figure 8. Trouble recorder card

tains the individual cards for the called area when 6-digit translation is required.

The IND1 hole in the second line and the IND2 hole in the fourth line are index holes and are never enlarged. Any card that drops will always cut off the light through those channels. This serves as an indication that a card has actually dropped and that the phototransistors associated with the other holes should be prepared for action. The index holes also aid in trouble detection and in proper disposition of calls where cards are deliberately omitted or where operators have dialed a blank code in error.

The class holes indicate such things as type of pulsing and nature of the signaling channel used on the trunk group out of the office.

The area code control holes in the third line are to tell the decoder what to do about dropping or spilling forward an area code registered in the sender or supplying an area code when none is registered. This information is needed primarily in connection with alternate routing. The alternate route pattern number holes tell the decoder at what point to enter its chain of alternate route relays for the first choice alternate. Provision is made for a maximum of 100 entry points.

The holes on the fourth line are for making proper dispositions of calls when no circuits are available on any routes, telling how many digits to expect on certain calls and other items of a detailed technical nature.

The code conversion holes on the fifth line supply the arbitrary digits to replace code digits on calls routed through step-by-step TO's or TC's. Provision is made for one, two, or three digits as required.

The variable spill control holes in the sixth line tell whether to spill all digits received, skip the first three code digits, or skip six code digits.

The remaining holes define the location on the equipment of the test leads for the trunk group over which the call will be routed.

The notches around the edges are used for positioning and removal of cards.

An individual card is removed from the stack by first keying the code to cause it to drop so that it may be identified. Since a card can be located easily in this manner, it is unnecessary to keep cards in any ordered position in the box.

At least one translator is provided in every decoder. It contains the cards for all offices in the home numbering area of the CSP, for certain operator codes, the single 3-digit card for each toll numbering area and a card for each toll line group out of the office that can be used as an alternate route. If there are other areas to which the volume of traffic is very high and for which 6-digit translation is required the cards for those areas are put in a second machine in each decoder. Cards for other areas are put in foreign area translators common to the office and accessible to all decoders on a one-at-a-time basis. An emergency translator is provided to permit removal of all cards to it from any translator which may require prolonged repair work.

TRAFFIC CONTROL PANEL

The traffic control panel is located in the operating room. The equipment in it consists of a key for each group used as an alternate route. When a particular key is operated, no alternate routed traffic will be offered to the group represented by it nor to any group above it in the fixed alternate route pattern. This is done to relieve offices which are overloaded by either unforeseen or predicted traffic peaks.

Maintenance

The maintenance facilities for the new CSP system are basically similar to those of the older toll crossbar system with the necessary addition of equipment to test the new features introduced. The sender test frame is, of course, obliged to test the CSP features added to the sender and the trouble indicator frame is changed to operate with the new decoders, translators, and markers.

In place of the lamp trouble indicator the new trouble recorder introduced with the latest local crossbar system[7] is used. Whenever trouble is encountered, it punches on a card a record of the circuits involved and of the important events that had occurred in the progress of the call, as an aid to the maintenance man in locating the trouble. A sample trouble recorder card is shown in Figure 8.

Automatic equipment for testing the operation and transmission features of intertoll trunk has been designed for the older systems and for the new CSP system.

Switching Aspects of Customer Toll Dialing

In the course of developing the switching system for CSP's, the requirement for handling long-distance traffic dialed by customers as well as that dialed by operators was kept in mind. The trial of long-distance customer dialing now in progress in Englewood, N. J. confirms the soundness of the basic plan and exemplifies the principles involved in full realization of the plan. With a toll network laid out to accept a distinctive 10-digit number for any telephone in the country and route it to the proper destination, the remaining tasks to be performed are to provide for delivery of the number to the toll network from the customer's dial instead of from an operator and to provide an automatic record of the call for charging purposes.

In Englewood both tasks were performed quite easily. The Englewood local office equipment is of the most modern type[7] and includes automatic message accounting (AMA)[8] facilities. When it was in the development stage, the ultimate requirement for nationwide customer dialing was foreseen and provision was made for expanding the digit capacity of the switching equipment at small expense. Also the designs of the accounting center were such that corresponding changes could be made readily. In the new local office switching system, arrangements were included for sending forward the complete number, as received, to the toll office by MF pulsing. The system also was designed to be capable of automatic alternate routing and this feature is used in the trial.

Expansion of the program will, of course, demand that similar arrangements be provided for the older types of local switching systems already in service More extensive modification will be required to make them capable of giving the customer the same service. For them, as for the most modern system, however, AMA equipment is admirable for recording the information necessary for charging for the calls.

The requirement for customer toll dialing that senders (or directors) and recording equipment be provided has a bearing on the type of equipment used at TC's and TO's. For calls handled by operators and for calls received by the customers through such offices the only disadvantage of step-by-step equipment without senders at those points is that the CSP equipment at other points must be somewhat more complicated and expensive than it would otherwise need to be. But with customer dialing, if senders and recording equipment are not provided either in the local office or in the TC or TO, the calls must be routed by the most direct means possible to a CSP where such equipment is provided. Thus some advantages that might be gained from having them at the TC or TO would be lost:

1. In some cases an indirect route to the CSP would need to be taken for the sole purpose of recording the call. For example, a call which normally might be switched from a TC through a TO to another TC would need to be connected to the CSP for making the record.

2. There is no operator at the TC or TO to select an alternate route and with the equipment there incapable of automatic alternate routing, the economies and service protection inherent in the alternate routing procedure would be lost.

If step-by-step toll switching equipment is already provided at a TC or TO, senders (or directors) could be added, making it in effect a common control switching system. This measure would permit automatic alternate routing and the further addition of recording equipment would eliminate the indirect routings for recording purposes.

A further benefit from having common control equipment in TC's or TO's can be realized in some instances. When a customer is served by a local office that has no senders, he must dial one or more directing digits (probably three digits) ahead of the 7- or 10-digit number in order to get to an office where senders are provided. It is, of course, desirable to avoid this extra burden on the customer. Where the equipment in a TC or TO can be used in common for switching local and toll traffic, the customers whose lines are terminated in that office will be dialing directly into senders, if the equipment uses common control, and will, therefore, benefit in that they will not have to dial directing digits.

Conclusion

The new system was designed to implement the nation-wide switching plan, which integrates the switching network of the entire nation into a single unit. This switching job, requiring a high order of mechanical intelligence, is the most comprehensive ever performed by any system. Skillful manipulation of code digits enables the provision of a numbering plan covering the entire country with a minimum number of digits to give each customer a distinctive number. It also obviates the need for extra expense to make step-by-step toll offices satisfactory operating elements of the plan in locations where CSP features are not essential.

The automatic and almost instantaneous selection of alternate routes makes it possible to give virtual no-delay service without greatly increasing the cost of outside plant and to make multiswitch connections at a speed comparable to that for local service.

The translating equipment simplifies administration of the plan which demands co-ordination of activities on a nation-wide basis.

The numbering plan, the switching plan, and the CSP equipment which implements them make it feasible to offer nation-wide dialing service to customers without the aid of operators when automatic charging facilities and local office switching arrangements for handling the three extra digits of the national number are provided. It will be readily appreciated that so far as the CSP switching equipment is concerned, it is immaterial whether the digits it receives come from an operator or from a customer. The call will be routed to its destination and supervision for charging purposes will be furnished in the same manner in either event.

The new system represents an important step in the process of continually improving the long-distance switching methods of the Bell System with consequent improvement of the service to all telephone customers in the United States and Canada.

References

1. Fundamental Plans for Toll Telephone Plant, J. J. Pilliod. *AIEE Transactions*, volume 71, part I, September 1952, pages 248–56.

2. Crossbar Toll Switching System, L. G. Abraham, A. J. Busch, F. F. Shipley. *Electrical Engineering* (*AIEE Transactions*), volume 63, June 1944, pages 302–09.

3. Application of Multifrequency Pulsing in Switching, C. A. Dahlbom, A. W. Horton, Jr., D. L. Moody. *AIEE Transactions*, volume 68, part I, 1949, pages 392–96.

4. Nationwide Numbering Plan, W. H. Nunn. *AIEE Transactions*, volume 71, part I, September 1952, pages 257–60.

5. Crossbar Dial Telephone Switching System, F. J. Scudder, J. N. Reynolds. *Electrical Engineering* (*AIEE Transactions*), volume 58, May 1939, pages 179–92.

6. Single-Frequency Signaling System for Supervision and Dialing over Long-Distance Telephone Trunks, N. A. Newell, A. Weaver. *AIEE Transactions*, volume 70, part I, 1951, pages 489–95.

7. The Number 5 Crossbar Dial Telephone Switching System, F. A. Korn, James G. Ferguson. *AIEE Transactions*, volume 69, part I, 1950, pages 244–54.

8. Fundamentals of the Automatic Telephone Message Accounting System, John Meszar. *AIEE Transactions*, volume 69, part I, 1950, pages 255–69.

No Discussion

Dynamic Hysteresis Loop Measuring Equipment

H. W. LORD
FELLOW AIEE

WITHIN the past 2 or 3 years many scientists and engineers working on research and development projects in the magnetic amplifier field have come to realize that the dynamic hysteresis loop of a magnetic material plays an important role in determining the amplification and control characteristics of high-gain magnetic amplifiers. Figure 1 shows a block diagram of an excitation and test circuit which often has been used to display the dynamic hysteresis loop of a core material upon the screen of a cathode-ray oscilloscope tube.

This circuit includes means for impressing upon the vertical, or *B*, axis of the oscilloscope a signal which is proportional to the flux density in the core and impressing upon the horizontal, or *H*, axis of the oscilloscope a signal which is proportional to the magnetizing force acting on the core specimen under test. The *B* axis channel includes a secondary (*B*) coil wound on the specimen under test, an integrating circuit, and a linear amplifier, the output of which supplies voltage to the vertical deflecting plates. The *H* axis channel consists of an amplifier to amplify the relatively low voltage drop across a resistor (current shunt) in series with the primary or exciting (*H*) coil so as to provide an adequate deflection voltage to the horizontal deflecting plates of the oscilloscope. The Lissajous figure shown by this system on the cathode-ray tube screen then will be indicative of the shape of the dynamic hysteresis loop of the specimen under test.

Figure 2 shows a 60-cycle dynamic hysteresis loop of Deltamax as displayed in the manner just described, by a good system of amplifiers and integrator circuits. Furthermore, by suitable calibrating means, the amplitude of the deflection in the vertical axis will indicate the average flux density in the core and the amplitude of the horizontal deflection will indicate the magnetomotive force applied to the core material.

On many occasions the hysteresis loop displayed by such a system has been found to show some very peculiar shapes. At such times one then begins to question the fidelity of the system being used to display the dynamic hysteresis loop.

This paper will endeavor to indicate the several sources of errors and their effects upon the displayed hysteresis loop. Also, it will indicate means for compensating for certain errors introduced by the amplifiers and will specify the general performance characteristics of amplifiers which will be satisfactory for displaying a dynamic hysteresis loop without introducing serious errors and/or distortions.

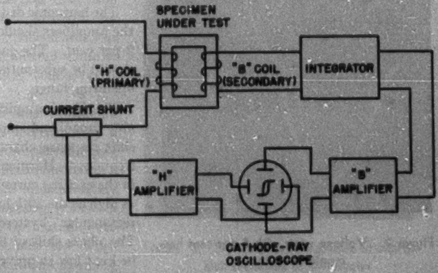


Figure 1. Block diagram of excitation and test circuit

Paper 52-238, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted February 4, 1952; made available for printing May 14, 1952.

H. W. Lord is with General Electric Research Laboratory, Schenectady, N. Y.

The author wishes to acknowledge the contributions of D. R. Kearns and R. Rescorla, who carried out the experimental phases of this work.

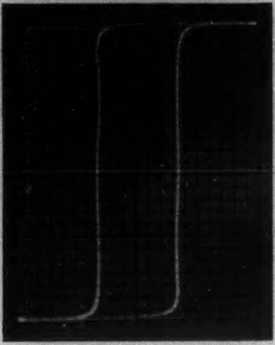


Figure 2. Dynamic hysteresis loop of 0.005-inch Deltamax at 60 cycles

Some Sources of Errors, Their Effects upon the Displayed Hysteresis Loop, and Means for Minimizing Them

To provide a readily reproducible method for testing cores in a manner approximating their use in magnetic amplifiers, a sinusoidal flux wave has been chosen. If the wave form of the average flux density in the core departs materially from that of a sine wave, the maximum rate of change of flux in the core will exceed that of a sine wave. Higher rates of change of flux will produce higher values of indicated dynamic coercive force and, therefore, will be in error. If the total harmonic distortion of the voltage induced by the flux wave is kept to less than 10 per cent, errors due to this effect will be low enough to be tolerable. Figure 3 shows an induced voltage wave with 6-per-cent total harmonic distortion due to saturation effects.

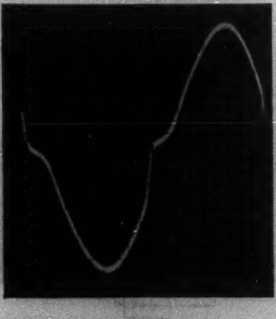


Figure 3. Voltage wave with 6-per-cent harmonic distortion

The source of power used to excite the specimen under test should be a good sine wave with low distortion. It should have a low internal impedance since it must supply highly distorted current wave shapes with only slight distortion of the voltage wave shape of the power source. The resistance of the exciting coil or primary should be kept low.

The current drawn by the exciting coils of the specimen often is highly distorted and may have very high ratios of peak to average values. The resistor used in series with the excitation circuit to act as a current shunt for supplying voltage to the *H* amplifier must be restricted to relatively low values of resistance. The peak voltage drop across this resistor, plus that across the resistance of the exciting coils, must be only a small fraction of the total voltage applied to the excitation circuit. Excessive nonsinusoidal voltage drops across these resistances will cause the voltage wave form impressed upon the excitation coil to be distorted seriously at peak flux densities which saturate the core.

The *B* coil, or secondary winding, should be fairly closely inductively coupled to the primary winding. However, care should be taken that the coupling capacitance between this winding and the excitation winding be kept low. This is particularly necessary when testing rectangular hysteresis loop core materials, since high-frequency components of voltage are present when testing under conditions of high saturation. Although these high-frequency components of voltage are small, excessive capacity coupling will show up as small oscillations superimposed upon the normal hysteresis loop.

The integrating circuit should have a high impedance in order that the effects of its loading upon the *B* coil be kept low. Excessive loading by this circuit appears as a widening of the hysteresis loop and therefore is another source of error in determining the true values for the dynamic coercive force and the true shape and area of the dynamic hysteresis loop.

The harmonic distortion contributed by the amplifiers should be held to less than 2 per cent. The gain versus signal level should be substantially constant over the operating range.

The *H* axis amplifier must be a reasonably broadband amplifier with low phase-shift response characteristics at the test frequency. Harmonic frequencies present in the exciting current may include values as high as the 20th harmonic when testing rectangular hysteresis loop materials. The phase shift of these harmonics must be kept low to prevent the portion of the hysteresis loop above the knee of the curve from appearing as shown in Figure 4. The phase shift of the amplifier at the testing frequency and below also should be kept low. Excessive low-frequency phase shift causes distortions as shown in Figure 5.

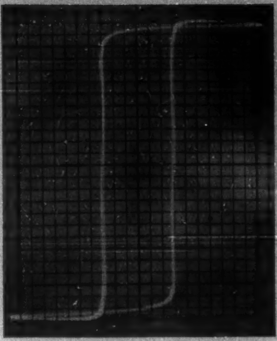


Figure 4. Distortion of loop due to excessive phase lag in H amplifier at high frequencies

Care should be taken to assure that this amplifier will not be overloaded by the high peak voltages present in the voltage drop across the *H* axis dropping resistor when testing to high saturations. The high ratio of peak-to-average of this voltage may overload this amplifier and cause errors in indication of the peak values of *H* under which the specimen is being tested. In some cases, clipping of this voltage is desirable to obtain accurate values of the dynamic coercive force under conditions of high peak values of

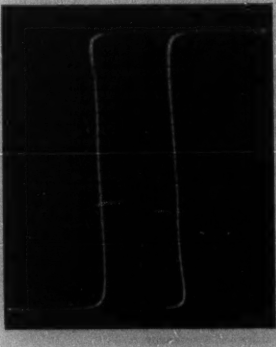


Figure 5. Distortion of loop due to excessive phase lead in H amplifier at low frequencies

magnetomotive force. However, any such clipping circuits should not effect the calibration or voltage gain of the amplifier within the desired range of linear response.

The voltage response characteristic of the *B* axis amplifier at frequencies above the test frequency is of secondary importance. However, the phase response characteristic at and below test frequency is very important. Excessive low-frequency phase shift (leading phase angle) causes a distortion of the loop of the type shown in Figure 6. Note that this is very similar in appearance to Figure 4 for the case in which the high-frequency phase shift of the *H* amplifier is excessive.

Figure 7 shows the effect of excessive high-frequency phase shift (lagging phase angle) of the combined *B* amplifier and integrator circuit. Note that the effect of this is to widen the loop tip instead of narrowing it or causing a reversed loop effect as shown by Figures 4 and 6.

If the frequency of the power source supplying the amplifier tube filaments and the plate supply rectifier circuit differs from that of the testing frequency, then hum effects are detected readily by a blurring or widening of the oscillograph trace over the entire figure being displayed by the oscilloscope. However, when the test frequency is synchronous with the source supplying power to the amplifier, the effects of hum in the amplifier are not so readily apparent. Hum effects from the amplifier filaments often cause symmetrical distortion of the hysteresis loop display. Its effect may be either a widening or a narrowing of some portion of the loop. The portion of the loop at which it appears depends upon the phase of the hum with respect to the excitation voltage. This type of hum effect usually can be detected by reversing the phase of the voltage applied to the excitation circuit and noting any change in the shape of the displayed figure.

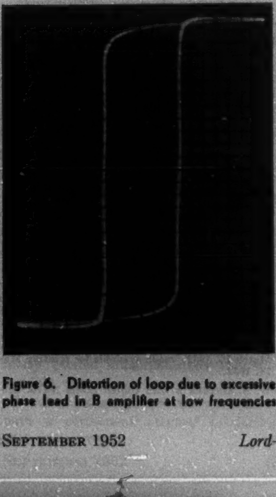


Figure 6. Distortion of loop due to excessive phase lead in B amplifier at low frequencies

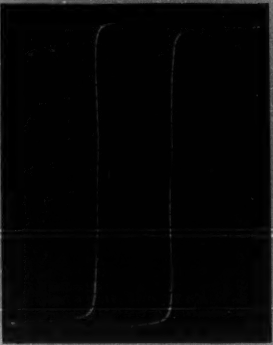


Figure 7. Distortion of loop due to excessive phase lag in B amplifier plus integrator at high frequencies

Hum due to insufficient filtering of the plate supply to the amplifier has a predominant second harmonic component. This causes an asymmetrical distortion of the hysteresis loop and is evidenced by a reduced deflection in one portion of the loop and increased deflection in the corresponding portion of the other half of the loop.

Criteria for Maximum Tolerable Errors

In order to specify the minimum performance requirements of measuring equipment such as this, it is necessary to set up some criteria upon which to judge whether or not the equipment is satisfactory. The criteria chosen for this dynamic hysteresis loop measuring equipment are as follows:

All amplitude or deflection distortion effects were judged as being satisfactorily low if the error they introduced was less than 5 per cent of the indicated value or less than one trace-width, whichever is greater.

The phase distortion was considered unsatisfactory if it was detectable by a close observation of the displayed figure shape. In many cases this amounted to a change corresponding to approximately the width of the oscillograph trace. A rectangular hysteresis loop material was used as the specimen under test when judging phase distortion effects upon the measuring circuit performance. This imposes severe requirements upon both the *B* and *H* channels. The *H* amplifier must handle a very wide frequency range and a phase distortion in the *B* amplifier is detected readily by a widening or a reverse looping of the trace above the knee of the hysteresis loop.

Characteristics of Amplifiers of Minimum Performance

The minimum performance characteristics required to meet the criteria specified will be indicated now. Each amplifier is assumed to be a 3-stage amplifier and is specified upon the basis that a good wide-band amplifier is used for the other axis. Somewhat poorer phase response characteristics could be tolerated if both amplifiers were identical and had a constant time delay. However, this possibility has been ignored in arriving at the requirements given here.

To keep the phase shift to a satisfactorily low value, each stage of the *H* axis amplifier should have less than 45 degrees of phase shift at a low frequency of 1/100th of test frequency and at a high frequency of 250 times test frequency. With conventional types of resistance-capacitance coupling networks and with screen and cathode circuits such that they do not contribute to the frequency response characteristics, the amplitude response per stage will be down 3 decibels at the frequencies just indicated.

The requirements for the *B* amplifier phase shift are very stringent on low-fre-

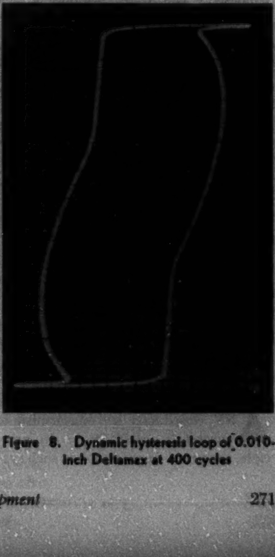


Figure 8. Dynamic hysteresis loop of 0.010-inch Deltamax at 400 cycles

quency response and a little less stringent than those of the *H* amplifier on the high-frequency end. If a single resistance-capacitance integrating circuit having a phase angle at operating frequency of 89.75 to 90 degrees is used, then the low-frequency phase response per amplifier stage must have a phase shift of less than 45 degrees at 1/300th of the operating frequency. However, if suitable compensation is provided by an additional corrective integrating network per amplifier stage, then the low-frequency phase response need be only less than 45 degrees at 1/30th of the operating frequency.

The high-frequency phase response of this *B* amplifier should be such as to have less than 45 degrees of phase shift in each stage at 100 times operating frequency.

With a suitable broad-band *B* amplifier, a simple single resistance-capacitance network of proper values will provide satisfactory integration. It should be such as to provide, across the capacitor, a voltage which lags the input voltage by 89.75 to 90 degrees at operating frequency.

When additional integrating types of resistance-capacitance networks are added to equalize for excessive low-frequency phase advance or lead due to the coupling networks in the amplifier, the equalizing networks should be such as to provide, at operating frequency, a total over-all lagging phase of 90 degrees ± 0.25 degree for the combined integrator and entire *B* axis amplifier circuit.

The maximum distortion in the voltage wave induced by the flux wave should not exceed 10 per cent. This includes distortion due to the power source and all series resistance in the excitation circuit.

As a final check for a complete system of these minimum performance specifications, networks representing this complete amplifier system were inserted between a test specimen and a broad-band oscilloscope. The loop displayed by this system was photographed and compared with photographs of the loop displayed when using a 90-degree integrator network and just the broad-band oscilloscope. The dynamic hysteresis loop displayed by the complete system met the original criteria set forth when, at test frequency, the integrator circuit was adjusted so that the phase of the output voltage of the *B* amplifier lagged the phase of the output voltage of the *H* amplifier by no less than 90 degrees nor more than 90.5 degrees.

The amplifier performance specified here is admittedly somewhat conservative. However, when one employs amplifiers meeting the specifications, along with carefully applied instrumentation, and obtains a dynamic hysteresis loop such as that shown by Figure 8, it is safe to assume that the displayed Lissajous figure truly represents the dynamic hysteresis loop of the specimen under the conditions imposed by the test circuit.

No Discussion

A New Type of Magnetic Servo Amplifier

W. A. GEYGER
MEMBER AIEE

Synopsis: This paper describes a new 2-stage magnetic amplifier for high-performance servomechanisms, particularly to be used in connection with those types of 2-phase induction motors having such a low ratio of effective viscous friction to inertia that it is difficult to stabilize at high gain in a closed loop. The basic idea consists in feeding the amplifier field winding of the a-c reversible motor by a variable unidirectional current having a fundamental frequency (60 or 400 cycle) component and a d-c component producing an effective dynamic braking of the motor, preferably in such a way that the d-c component increases when the a-c component decreases, and vice versa.

There are various possibilities for designing of magnetic servo amplifier circuits which are based upon the aforementioned principle. However, only two typical 2-stage circuit arrangements will be treated in this paper to show the unique properties of these circuits which have inherent dynamic braking characteristics.

ANY closed-loop servomechanism is inherently oscillatory. In principle, the use of amplification in such systems is seldom possible without introducing a certain time-lag between cause and effect (due, for example, to the inductance of a field coil arrangement), which gives rise to hunting. The reason for the instability is that the correcting means persists a little too long, hence the error is overcorrected. A reversal of the correcting means follows, which again overshoots, and, unless some antihunt method is employed for cancelling the tendency to swing, it will continue indefinitely.

Unfortunately, the use of magnetic amplification is associated with a certain additional time-lag of the magnetic amplifier itself. However, it should be born in mind that the lags arising from external parts of the servomechanism are often appreciably greater than those inherent in the magnetic amplifier. For example, when working with a power supply frequency of 400 cycles per second, the usual response time of about 6 to 20 cycles (15 to 50 milliseconds) represents a performance which will not cause any difficulty in many cases. Of course, if the magnetic servo amplifier is designed to operate from a power supply of 60 cycles per second, a response time of 6 to 20 cycles (100 to 333 milliseconds) may not be acceptable. Thus, in this case, a very short response time of about 1 cycle only is really desirable.

F. G. Logan[1] has shown that this may be accomplished in various ways by using a-c controlled self-saturating circuits, the alternating control current having preferably the same frequency as that of the power supply generator. Other types of self-saturating circuits having a response time of 1 cycle only have been described by R. A. Ramey[2,3] and A. E. Schmid.[4] An inherent characteristic of such circuits is that control is achieved in full during each cycle and that when the controlling current is changed, full response to this change is reached not longer than 1 cycle later. Of course, application of these self-saturating circuits having inherent 1-cycle response will be particularly valuable for the design of magnetic servo amplifiers to be operated from a 60-cycle-per-second power supply.

Paper 52-237, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted March 10, 1952; made available for printing May 14, 1952.

W. A. Geyger is with United States Naval Ordnance Laboratory, Silver Spring, Md.

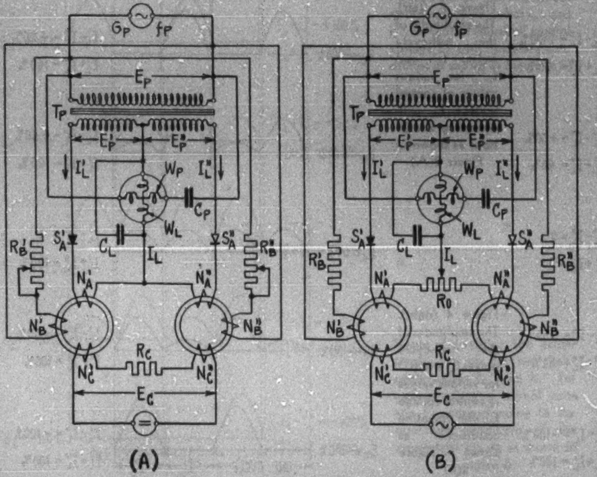


Figure 1. Two types of full-wave push-pull output-stage circuits having inherent dynamic braking characteristics with d-c control windings (A) or with a-c control windings (B)

Common remedies for hunting are damping the motor itself by introducing dynamic braking, application of a resistance-capacity differentiating network or tachometer to stabilize by means of derivative feedback, and magnetic or viscous flywheel dampers.[5] The purpose of such devices is to provide damping at a rate appropriate to the requirements of stability.

Another important factor governing the satisfactory operation of a closed-loop servomechanism is the proper adjustment of the sensitivity or gain of the magnetic amplifier. The combination of the error signal voltage per degree and the actual amplifier gain governs the tightness of the system. If the gain is too great, then the servomechanism will oscillate in spite of application of an effective antihunt method. It is often desirable to adjust the gain of the magnetic servo amplifier in such a way that not more than one overshoot, or no overshoot at all, will be obtained.

Referring to the damping properties of 2-phase induction motors which are used extensively in high-performance servomechanisms, it is important to stress that some of these motors have good viscous friction to inertia ratios or special damping means,[6] and others have not. Separately excited induction-meter type reversible motors, as sometimes used in self-balancing potentiometer recorders,[6] have a permanent magnet[6] producing eddy-current damping. Some types of low-inertia squirrel-cage induction motors and drag-cup type motors have eddy-current damping devices built into the motor. There are, however, other constructions of such motors which have virtually no damping properties so that additional damping methods must be applied to prevent hunting.

The use of vacuum-tube amplification offers the possibility of introducing effective dynamic braking of the motor by means of the d-c component flowing through the plate circuit, which contains the amplifier field winding of the 2-phase induction motor. However, if such a motor is to be controlled by means of a magnetic servo amplifier, considerable practical difficulty will be caused by the fact that the alternating load current of the conventional magnetic amplifiers of the balance detector type[6] does not have a d-c component which may be used for producing dynamic braking of the motor.

This paper describes a new type of 2-stage magnetic servo amplifier, for use particularly with 2-phase induction motors having very poor damping properties. The basic idea consists in feeding the amplifier field winding of the a-c reversible motor by a variable unidirectional current having a fundamental frequency (60 or 400 cycle) component and a d-c component producing an effective dynamic braking of the motor, preferably in such a way that the d-c component increases when the a-c component decreases, and vice versa.

There are various possibilities for designing of 2-stage magnetic servo amplifier circuits which are based upon the aforementioned principle. However, only two typical circuit arrangements will be

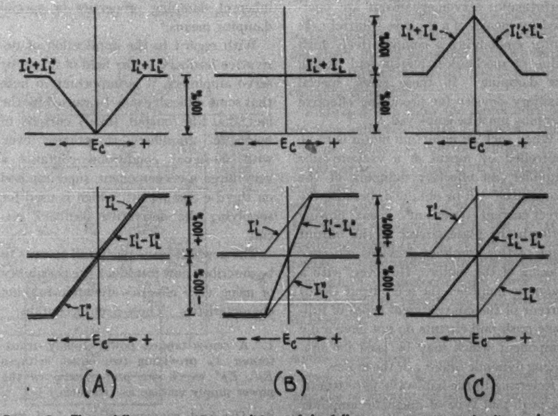


Figure 2. Three different operating conditions of the full-wave output-stage circuits, as shown in Figure 1

(A). There is no damping effect with no-signal conditions
(B). There is a constant damping effect
(C). There is a variable damping effect

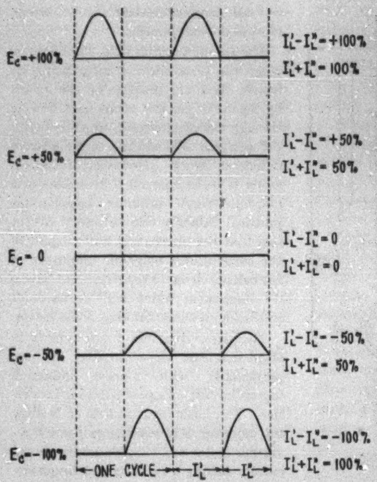


Figure 3 (left). The amplitudes of the half-cycle pulses I_L' and I_L'' for various values of control voltage E_C with operating conditions as shown in Figure 2(A)

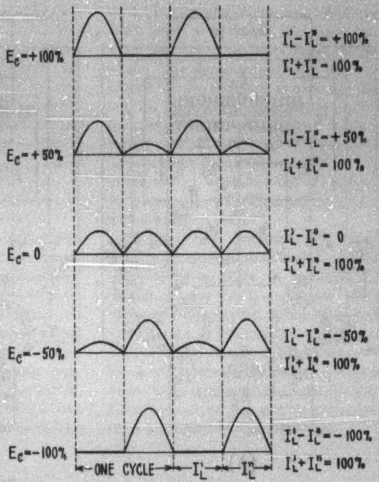


Figure 4 (right). The amplitudes of the half-cycle pulses I_L' and I_L'' for various values of control voltage E_C with operating conditions as shown in Figure 2(B)

treated in this paper to show the unique properties of these circuits which have inherent dynamic braking characteristics.

Damping Requirements

The conventional magnetic amplifiers of the balance detector type[6] are particularly suitable for operating those constructions of reversible induction motors which have inherent damping or special damping means.[5] However, some high-performance servomechanisms are operated by means of a 2-phase squirrel-cage motor embodying a comparatively high value of mechanical inertia and having low damping. In these cases, special auxiliary devices for producing effective damping must be employed.

When such an induction motor is to be controlled by means of a vacuum-tube amplifier, an effective damping of the motor itself will be obtained, because the direct current (quiescent current) flowing through the plate circuit of the output stage of this amplifier provides dynamic braking of the motor. However, with a magnetic amplifier the a-c output (load) current of the conventional types of full-wave push-pull circuits do not have a d-c component which may be used for producing dynamic braking of the motor.

Referring to the half-wave self-saturating push-pull type of magnetic amplifier a-c motor control circuits, as described by C. S. Hudson,[7] L. R. Crow,[8] and others, it is to be noted that the torque-producing a-c component and the associated d-c component of the unidirectional reversible motor control current will always increase or decrease at the same rate. Thus, there is no damping effect at all with no-signal conditions. This is generally not desirable when operating with those constructions of 2-phase induction motors which have very poor damping properties because there is no dynamic braking effect in the vicinity of the null point. However, these half-wave push-pull type circuits have proved to be very useful with a reversible a-c motor having inherent damping properties or special damping means.[5]

With regard to the application of derivative feedback in the field of magnetic servo amplifiers, it is important to note that some difficulty may be caused by the fact that the output (load) current of magnetic amplifiers—sometimes even with no-signal conditions—contains a very large a-c component superimposed on the d-c component, which is used for supplying the derivative feedback circuit.

The new full-wave push-pull circuits to be described now introduce the possibility of using very effective dynamic braking characteristics. These circuits contain:

1. A center-tapped power supply transformer T_P providing two equal voltages E_P', E_P'' which are proportional to the power supply voltage E_P (frequency f_P).
2. The load windings N_A', N_A'' of two equally rated saturable reactor elements, preferably having Orthonol tape cores.[9,10]
3. Two series-connected half-wave selenium rectifier sets S_A', S_A'' producing self-saturating effects.
4. A-c bias windings N_B', N_B'' with series resistors R_B', R_B''.
5. D-c control windings, Figure 1(A), or a-c control windings, Figure 1(B), N_C', N_C'' with a series resistor R_C.
6. The amplifier field winding W_L of the 2-phase reversible induction motor with shunt capacitor C_L.
7. The line field winding W_P of this motor with series capacitor C_P.

The basic full-wave output-stage circuits (Figure 1) differ from half-wave push-pull circuits[7,8] in that positive half-waves of current (I_L') flow through load winding N_A', and negative half-waves of current (I_L'') flow through load winding N_A''. In any case, the two control windings N_C' and N_C'', which may be connected with various types of d-c or a-c control circuits, are arranged so that push-pull action is obtained.

The a-c bias circuits $N_B'R_B'$ and $N_B''R_B''$ are provided to eliminate differences of the magnetic properties of the two saturable reactor elements, to bring the operating point of these elements to its most favorable position, and to produce special voltage-current characteristics of the magnetic amplifier. Another way for eliminating such differences consists in using a potentiometer-type resistor R_O to be adjusted so that the reversible motor will not move, if the control voltage E_C is zero.

Figure 2 illustrates three different operating conditions of the full-wave output-stage circuits of Figure 1 having inherent dynamic braking characteristics.

Figure 2(A): With $E_C = 0$ the two load

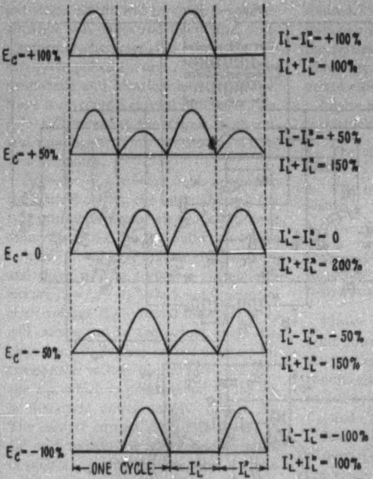


Figure 5 (left). The amplitudes of the half-cycle pulses I_L' and I_L'' for various values of control voltage E_C with operating conditions as shown in Figure 2(C)

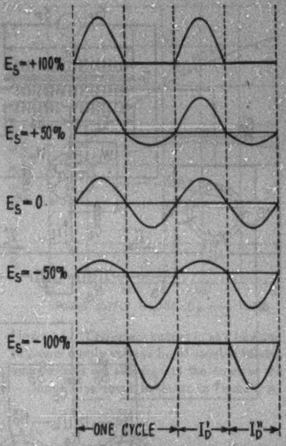


Figure 7 (right). The actual amplitudes of the two half-cycle pulses I_D' and I_D'' for various values of control voltage E_S, corresponding to the operating conditions of the input-stage circuit as shown in Figure 6

(quiescent) currents I_L' and I_L'' are equal and very small, and there is practically no damping effect because the d-c component of $I_L'+I_L''$ is extremely small. However, $I_L'+I_L''$ and the corresponding damping effect increases, when E_C increases. This is generally not desirable, because the speed of the motor is decreased with higher values of error voltage E_C and because there is practically no damping effect in the vicinity of the null point.

Figure 2(B): With $E_C=0$ the two load (quiescent) currents I_L' and I_L'' are about 50 per cent of maximum load current, and there will be a large and constant damping effect because $I_L'+I_L''$ will have a constant magnitude over the entire range. These working conditions will give satisfactory results in many cases.

Figure 2(C): With $E_C=0$ the two load (quiescent) currents I_L' and I_L'' have about their maximum (100 per cent) values, and there will be a very high damping effect. It is to be noted, however, that the damping effect decreases if the error voltage E_C increases. This variable damping effect which is produced by the changing of the d-c component $I_L'+I_L''$ permits faster motor movement toward the null point and slower movement coincident with the lower values of error voltage E_C characterizing the close proximity of the null point. Thus, a very effective dynamic braking of the motor will be obtained in this way: When the a-c (60 or 400 cycle) component of load current $I_L=I_L'-I_L''$ disappears from the amplifier field winding W_L, the comparatively large d-c component of I_L tends to lock the rotor in the position assumed at the instant a-c excitation ceases.

Of course, it is possible in special cases to superimpose an additional d-c damping effect in the amplifier field winding (W_L) for further modification of the actual damping conditions.

Figures 3 to 5 show the amplitudes of the two half-cycle pulses I_L' and I_L'' for various values of control voltage E_C: +100 per cent, +50 per cent, 0, −50 per cent, and −100 per cent. The corresponding values $I_L'-I_L''$ of the a-c (60 or 400 cycle) component and the corresponding values $I_L'+I_L''$ of the d-c component also are indicated in this presentation which characterizes the three different operating conditions, as illustrated in Figure 2.

It is to be noted, however, that sinusoidal wave shape of the two half-cycle pulses I_L' and I_L'' has been assumed for purpose of illustrating the basic principle of push-pull full-wave output-stage circuits, as shown in Figure 1. The fact that the wave shape of saturable reactor type circuits is not sinusoidal does not change the operation principle of these circuits.

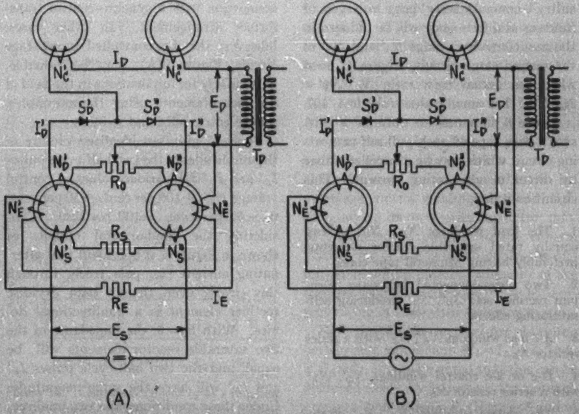


Figure 6. A new type of full-wave push-pull input-stage circuit containing two saturable reactor elements only, with d-c control windings (A) or with a-c control windings (B)

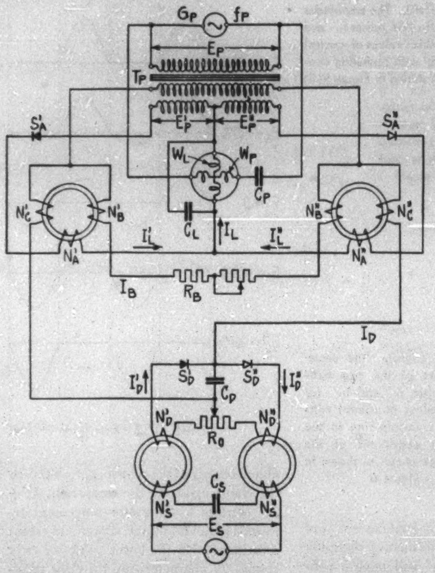


Figure 8. The new type of full-wave 2-stage magnetic servo amplifier circuit having a center-tapped power supply transformer (T_P)

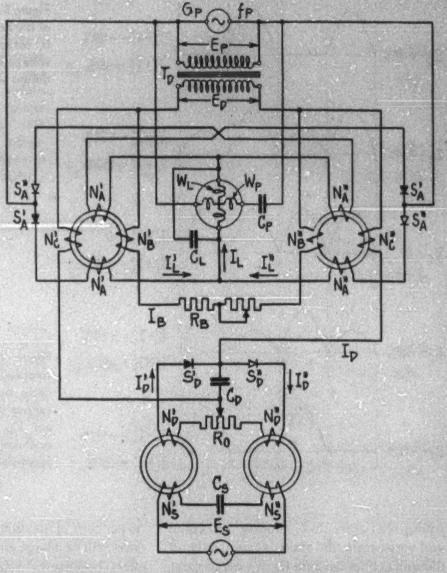


Figure 9. The new bridge type of full-wave 2-stage magnetic servo amplifier circuit having no power supply transformer

Basic Full-Wave Input-Stage Circuits

The various purposes of the input-stage circuit to be combined with the output-stage circuit, Figure 1, can be defined as follows:

1. To act as a push-pull type magnetic amplifier to produce a variable current, a function of the reversible error voltage E_S, for controlling the output-stage circuit.
2. To act as a phase discriminator (in a way similar to a phase-sensitive rectifier) so that actual direction of the resultant output current component $I_D=I_D'-I_D''$ will correspond to the actual phase displacement (0 degrees or 180 degrees) between error voltage E_S and power supply voltage E_P.
3. To act as a voltage-sensitive bias element to the output-stage circuit for introducing special voltage-current characteristics.
4. To act as a decoupling device in such a way that no appreciable alternating voltage will be induced in the synchro-transformer a-c control circuit having a total impedance of about 10,000 ohms.

The designer, when outlining the circuitry to be used in this special case, faces the following problem: The output currents of the push-pull input-stage circuit, preferably having two saturable reactor elements only, must provide

1. A bias component which is a function of the actual magnitude of power supply voltage E_P.
2. A reversible control component, which is a function of the actual error voltage E_S.

Furthermore, it is to be noted that—in contrast to conventional d-c control circuitry—comparatively large voltages of fundamental frequency will be induced in the two control windings N_C' and N_C'' of the output-stage circuit, Figure 1, even when the actual turn ratio $N_C'/N_A'=N_C''/N_A''$ has small value (about 1/40).

Figure 6 shows as an example a suitable 2-core type of push-pull self-saturating circuit which may be controlled either by direct or alternating current. This circuit contains:

1. The load windings N_D', N_D'' of two equally rated saturable reactor elements, preferably having Mumetal tape cores.
2. Two series-connected half-wave selenium rectifier sets S_D', S_D'' producing self-saturating effects.
3. A-c bias windings N_B', N_B'' with a series resistor R_B.
4. D-c or a-c control windings N_S', N_S'' with a series resistor R_S.
5. The control windings N_C , N_C'' of the push-pull output-stage circuit, as shown in Figure 1(A).
6. A potentiometer-type resistor R_O for zero adjustment.
7. A small transformer T_D (for example, 115/20 volts).

The a-c controlled input-stage circuit, Figure 6(B), will be used in the 2-stage magnetic servo amplifier operating in connection with a synchro-control transformer arrangement. In other cases, however, the d-c controlled input-stage circuit, Figure 6(A), may be valuable, particularly for applications in the field of d-c measurements using thermocouples, photoelectric cells, and so forth.

Figure 7 gives an idealized picture of the amplitudes of the two half-cycle pulses I_D' and I_D'' for various values of control voltage E_S: +100 per cent, +50 per cent, 0, −50 per cent, −100 per cent. Considering the fundamental circuit, as shown in Figure 6, it is evident that alternating current can pass freely through this circuit even though each dry-disc rectifier element is a unidirectional device. With $E_S=0$, the impedance of the two saturable reactor elements will be equal, and the two half-cycle pulses I_D' and I_D'' will have the same magnitude. Under these conditions, the two unidirectional currents I_D', I_D'' flowing through the control windings N_C', N_C'' of the output-stage circuit will produce a voltage-sensitive bias effect and the desired voltage current characteristics. Since the two half-cycle pulses I_D' and I_D'' are equal in magnitude and identical in wave shape, there is no control current difference $I_D = I_D' - I_D''$ present in the output-stage control windings N_C', N_C'' to run the a-c reversible motor.

However, if a voltage E_S is applied to the control circuit, the impedances of the two saturable reactor elements will have different values, the two half-cycle pulses I_D' and I_D'' will have different magnitudes, and there will be a control current difference $I_D = I_D' - I_D''$ which is a function of error voltage E_S. Actual direction of I_D will correspond to the actual phase displacement (0 degrees or 180 degrees) between error voltage E_S and reference voltage E_D, which is proportional to the 60- or 400-cycle power supply voltage E_P. The control current difference $I_D = I_D' - I_D''$ will operate the a-c reversible motor in such a way that the direction of movement will correspond to the actual phase displacement (0 degrees or 180 degrees) between the voltages E_S and E_D.

Full-Wave 2-Stage Circuits

Correct and efficient operation of the complete 2-stage circuit will be obtained only when the a-c (60 or 400 cycle) components of the currents flowing through the various parts of this circuit have a proper phase relationship. The following two examples show a simple way for obtaining proper phase displacement by means of four capacitors so that the most favorable working conditions will be secured. The final 2-stage circuits are results of a special experimental study, and the optimum values of the capacitors, which are not critical, have been determined empirically.

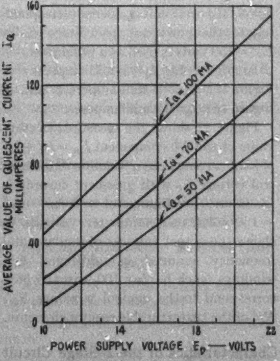


Figure 10. Quiescent current I_Q as a function of power supply voltage E_D, if constant biasing ampere-turns are supplied to the cores of the input-stage circuit as shown in Figure 6

Figure 8 shows a full-wave 2-stage circuit which contains:

1. The output-stage circuit, as shown in Figure 1.
2. The input-stage circuit, as shown in Figure 6.
3. The power supply transformer T_P having a center-tapped secondary winding (voltages E_P', E_P'') and an additional secondary winding supplying the input-stage circuit.

Proper phase relationship of the currents flowing through the various parts of this 2-stage circuit and maximum torque conditions of the 2-phase motor will be obtained by means of the capacitors C_P, C_L, C_D, and C_S.

It is to be noted that the actual values of output-stage quiescent currents I_L', I_L'' can easily be controlled by adjusting the variable a-c bias resistor R_B.

The bridge-type output-stage circuit, Figure 9, makes it possible to eliminate the power supply transformer (T_P in Figure 8) and to reduce size and weight of the complete magnetic servo amplifier by dividing the load windings N_A', N_A'' in two equally rated parts which are insulated from each other. It is to be noted that the total number of dry-disc rectifier cells remains unchanged, so that, for example, 4 by 12 cells may be used in the bridge type circuit (Figure 9), whereas 2 by 24 cells have to be provided in the transformer type circuit (Figure 8).

It is to be noted that application of the small power supply transformer T_D (for example, 115/20 volts) makes it possible to reduce the number of cells of the rectifiers S_D', S_D'' and to reduce the number of turns of input-stage load windings N_D', N_D'' and output-stage control windings N_C', N_C''. The transformer T_D, however, also may be eliminated by properly varying these circuit components.

Furthermore, in any case, the two input-stage control windings N_S' and N_S'', each having its own series capacitor, may be connected in parallel to the control voltage E_S of the synchro control transformer to obtain perfect symmetry of the transfer characteristic $I_L = f(E_S)$ by correct rating of these two series capacitors.

Referring to the method for obtaining perfect zero adjustment ($I_L = I_L' - I_L'' = 0$, if $E_S = 0$), it is to be noted that this can be achieved easily by adjusting the potentiometer—type resistor R_p (Figures 8 and 9) which controls the actual value of current ratio I_D'/I_D'' of the input-stage circuit.

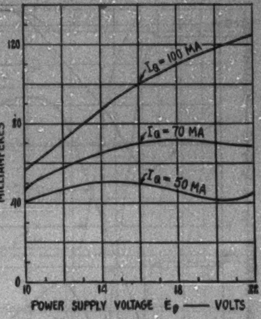


Figure 11. Quiescent current I_Q as a function of power supply voltage E_D, if variable biasing ampere-turns are supplied to the cores of the input-stage circuit as shown in Figure 6

Results of Measurements

To get some information regarding the actual performance of the new full-wave push-pull type circuits, various measurements on a representative 2-stage design have been made. The results of these measurements show that this magnetic servo amplifier has unique performance characteristics when nickel-iron-alloy tape cores with multilayer toroidal windings are used in the actual circuits.

Characteristics of the Input-Stage Circuit

This circuit, Figure 6(B), contains two Mumetal 2-mil tape cores (outside diameter: 2 inches, inside diameter: $1^3/_8$ inches, tape width: 0.5 inch) having multilayer toroidal windings ($N_D = 2{,}000$, $N_B = 50$, $N_S = 200$, each with number 28 wire size). The magnetic circuit has an effective area of 0.084 square inch = 0.54 square centimeter, and a mean length of 5.7 inches = 14.5 centimeters.

Selenium rectifier sets are used in such a way that the maximum value of reverse voltage on each rectifier cell does not exceed about 10 volts, to obtain favorable working conditions regarding the reverse resistances and the corresponding reverse-current components of the self-saturating circuit.

Referring to bias circuitry for magnetic amplifiers, it is important to stress that there are two ways of injecting into the core a magnetic biasing flux:

1. By supplying the core with constant biasing ampere-turns.
2. By supplying the core with variable biasing ampere-turns, preferably being pro-

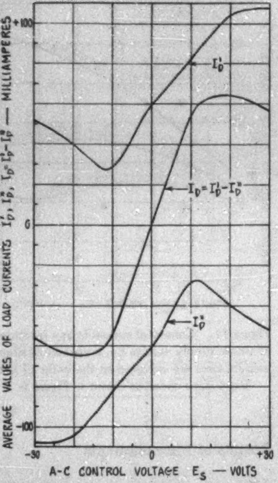


Figure 12. Load currents I_D' and I_D'' and control current difference $I_D = I_D' - I_D''$ as a function of the alternating control voltage E_S when the input-stage circuit, Figure 6(B), is used

portional to the power supply voltage of the magnetic amplifier.

Ten years ago the author[11] suggested designing magnetic servo amplifiers utilizing external feedback or self-saturation in such a way that the magnitude of the quiescent current is substantially independent of changes in a-c supply voltage. This was accomplished by applying special voltage-sensitive bias circuits which introduce inherent constant-current characteristics in the magnetic amplifier.

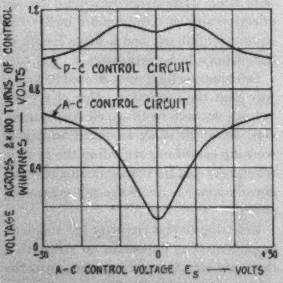


Figure 13. Actual voltage of fundamental frequency which is induced in series-connected control windings N_S' and N_S'', both for a d-c control circuit, Figure 6(A), and for an a-c control circuit, Figure 6(B)

Figure 10 shows the quiescent current I_Q as a function of power supply voltage E_D, if constant biasing ampere-turns are supplied to the cores of the input-stage circuit, as shown in Figure 6(B). The actual quiescent current value $I_Q = 50$, 70, or 100 milliamperes corresponding to the power supply voltage $E_D = 16$ volts is used as a parameter. Figure 11 shows the voltage-current characteristic $I_Q = f(E_D)$ of the same circuit, if variable biasing ampere-turns being proportional to E_D are used.

The results of these measurements using either constant or voltage-sensitive bias circuitary, as illustrated in Figures 10 and 11, show in a dramatic way the progress which is of paramount importance for successful development of magnetic servo amplifiers which must have an extremely small drift rate, so that the performance of the servomechanism is not appreciably influenced by changes of from ±5 to 10 per cent in power-supply voltage. It is interesting to note that special voltage-current characteristics, as shown in Figure 11, are due to the inherent negative-feedback properties of voltage-sensitive bias circuits: When the power-supply voltage E_D increases, then the bias current I_B (Figure 6) increases correspondingly so that the quiescent current I_Q will be controlled as to have a constant magnitude. Thus, the bias windings N_B', N_B'' carrying $I_B = \text{const} \times E_D$ act as additional control winding,s to make I_Q practically independent of changes within the limits from ±5 to 10 per cent in voltage E_D.

Figure 12 shows the load currents I_D', I_D'' and $I_D = I_D' - I_D''$ as a function of the alternating control voltage E_S when the a-c controlled input-stage circuit, Figure 6(B), is used. Figure 13 shows the actual voltage of fundamental frequency which is induced in series-connected windings N_S', N_S'' of a d-c control circuit, Figure 6(A), and an a-c control circuit, Figure 6(B).

Characteristics of the Output-Stage Circuit

This circuit (Figure 1) contains two Orthonol 2-mil tape cores (outside diameter: 2 inches, inside diameter: $1^5/_8$ inches, tape width: 1.0 inch) having multilayer toroidal windings ($N_A = 4{,}000$, $N_B = 100$, $N_C = 100$, each with number 31 wire size). The magnetic circuit of each core has an effective area of 0.168 square inch = 1.08 square centimeter, and a mean length of 5.7 inches = 14.5 centimeters.

Two selenium rectifier sets (each having 24 cells) are used in connection with a center-tapped power supply transformer T_P providing two equal secondary voltages $E_P' = E_P'' = 150$ volts, to obtain favorable working conditions regarding the reverse resistances and the corresponding reverse-current components of the self-saturating circuit.

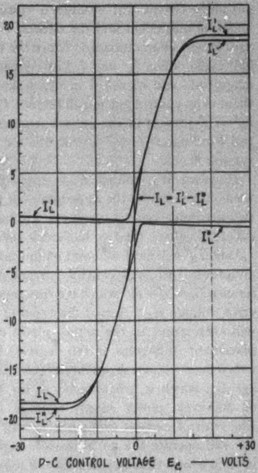


Figure 14. Load currents I_L', I_L'', and $I_L = I_L' - I_L''$ as a function of d-c control voltage E_C with actual quiescent current value $I_Q = 3$ milliamperes

Figures 14 to 16 show the measured values of load currents I_L', I_L'', and $I_L = I_L' - I_L''$ as a function of d-c control voltage E_C with $R_C = 1{,}000$ ohms; Figure 1 (A). The actual quiescent current value $I_Q = 3$, 10, or 17 milliamperes corresponding to the power-supply voltage $E_P' = E_P'' = 150$ volts is used as a parameter.

In Figure 14, $I_Q = 3$ milliamperes; in Figure 15, $I_Q = 10$ milliamperes; and in Figure 16: $I_Q = 17$ milliamperes.

Figure 17 shows the measured average value of the d-c component $I_{DC} = I_L' + I_L''$ of load current I_L as a function of d-c control voltage E_C with quiescent current I_Q (3, 10, or 17 milliamperes with $E_P' = E_P'' = 150$ volts) as a parameter. Figure 18 shows the actual voltages of fundamental frequency induced in additional test windings (each having 100 turns) which correspond to the control windings N_C', N_C'' of the two saturable reactor elements.

Characteristics of the 2-Stage Circuit

This circuit (Figure 9) has been investigated in connection with a standard type of servomechanism which is operated

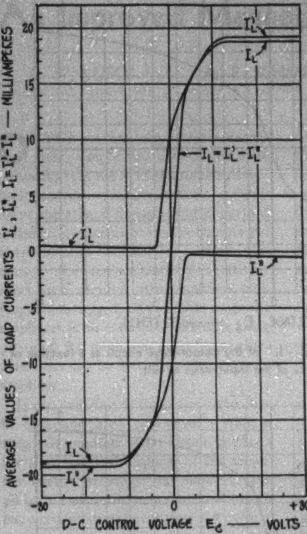


Figure 15 (left). Load currents I_L', I_L'', and $I_L=I_L'-I_L''$ as a function of d-c control voltage E_C with actual quiescent current value $I_Q=10$ milliamperes

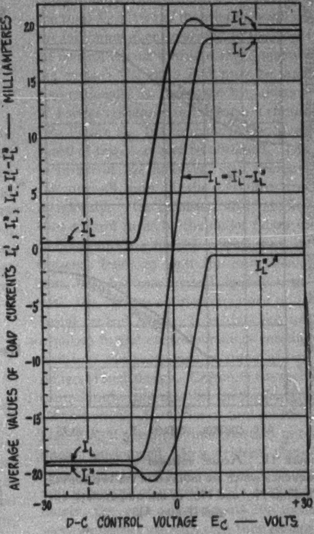


Figure 16 (right). Load currents I_L', I_L'', and $I_L=I_L'-I_L''$ as a function of d-c control voltage E_C with actual quiescent current $I_Q=17$ milliamperes

by means of a 3.5-watt 2-phase reversible induction motor having very poor damping properties and no additional damping means.

Figure 19 shows the resultant a-c load current $I_L=I_L'-I_L''$ of the output-stage circuit as a function of the a-c control voltage E_s of the input-stage circuit. Figure 20 shows the actual rotor speed of the 2-phase motor under no-load conditions as a function of the a-c control voltage E_s of the input-stage circuit. Referring to Figures 19 and 20, it is evident that an error voltage of about ±10 volts will give full output power of the magnetic amplifier.

As a result of various measurements it was learned that the static error is about 0.05 degree, while the drift error is about 0.1 to 0.2 degree.

Special Problems

A comparison between the conventional type 2-stage circuits having generally

1. eight saturable reactor elements (four elements in each stage),
2. four pairs of dry-disc rectifier elements (two pairs in each stage), and
3. an additional phase-sensitive rectifier circuit

and the new type of 2-stage circuits (Figures 8 and 9) having

1. four saturable reactor elements only (two elements in each stage),
2. two pairs of dry-disc rectifier elements only (one pair in each stage), and
3. no additional phase-sensitive rectifier circuit

shows that considerable economy will be achieved by the use of half the amount of basic components (saturable reactors and dry-disc rectifier elements), particularly with respect to ease of manufacture.

Referring to some special problems, which are involved in magnetic amplifier push-pull circuitry using two saturable reactor elements in each stage only, it is interesting to stress the following facts:

1. Those d-c controlled magnetic amplifier push-pull circuits, containing four saturable reactor elements in each stage,[4] have the unique feature that the fundamental frequency voltages, which are induced in the four series-connected control windings, are always cancelling each other completely, not only with no-signal

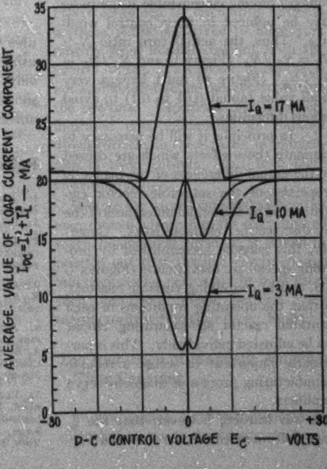


Figure 17. Average value of the d-c component $I_{DC}=I_L'+I_L''$ of load current I_L as a function of d-c control voltage E_C with quiescent current value I_Q as a parameter

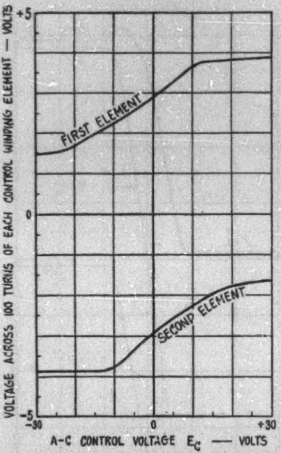


Figure 18. Actual voltages of fundamental frequency which are induced in the two control windings N_C' and N_C'' of the output-stage circuit (Figure 1)

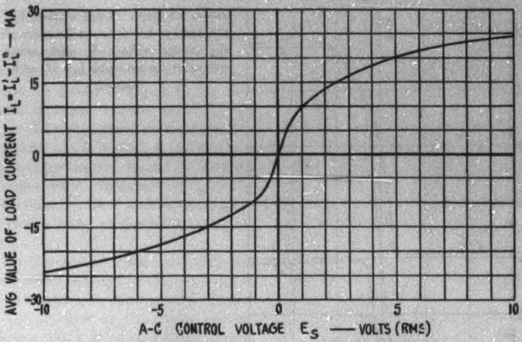


Figure 19. Resultant a-c load current $I_L = I_L' - I_L''$ of the output-stage circuit as a function of the a-c control voltage E_S of the input-stage circuit

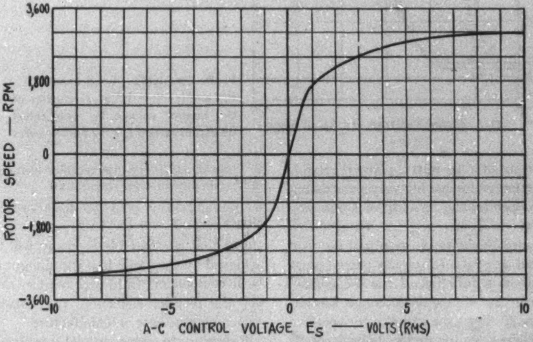


Figure 20. Actual rotor speed of the 2-phase induction motor under no-load conditions as a function of the a-c control voltage E_S of the input-stage circuit

conditions, but even when the maximum value of control voltage is applied to the amplifier.

However, any magnetic amplifier push-pull circuit containing two saturable reactor elements in each stage only, as described in this paper, has the typical property that the fundamental frequency voltages, which are induced in the two series-connected control windings, will cancel each other for no-signal conditions only; otherwise fundamental frequency voltages having considerable magnitudes may be induced in these control windings. Thus, the actual turn ratio $N_C'/N_A' = N_C''/N_A''$ (Figure 1) and $N_S'/N_D' = N_S''/N_D''$ (Figure 6) must have a very small value (about 0.01 to 0.1) to avoid serious difficulties.

2. In principle, it will be necessary to eliminate those effects which are caused by the actual differences of the magnetic properties of the saturable reactor elements, and by the actual differences of the dry-disc rectifier characteristics. Generally, this may be accomplished by providing special a-c bias circuits (Figures 1, 6, 8, and 9) containing variable resistors, so that the operating conditions of each saturable reactor self-saturating circuit can be adjusted individually. This is particularly important to secure a reliable manufacturing process of magnetic servo amplifiers.

It was learned, however, that the 2-stage full-wave circuit, as shown in Figures 8 and 9, can be adjusted easily by using one potentiometer resistor (R_o) only. This fact means a considerable simplification of the complete circuit for universal application.

References

1. Electric Controlling Apparatus, **F. G. Logan.** United States Patent *2,259,647*, October 21, 1941.

2. On the Mechanics of Magnetic Amplifier Operation, **R. A. Ramey.** *AIEE Transactions*, volume 70, part II, 1951, pages 1214–22.

3. On the Control of Magnetic Amplifiers, **R. A. Ramey.** *AIEE Transactions*, volume 70, part II, 1951, pages 2124–28.

4. Application of a 60 Cycle Magnetic Amplifier to a Position Indicating Servomechanism, **A. E. Schmid.** *Proceedings*, National Electronics Conference (Chicago, Ill.), volume 7, 1951, pages 515–22.

5. Electronic Instruments (book), **I. A. Greenwood, J. V. Holdam, D. Macrae.** Massachusetts Institute of Technology Radiation Laboratory Series, McGraw-Hill Book Company, Inc., New York, N. Y., chapter 11, 1948, pages 319–61.

6. Magnetic Amplifiers of the Balance Detector Type—Their Basic Principles, Characteristics, and Applications, **W. A. Geyger.** *AIEE Transactions*, volume 70, part II, 1951, pages 1707–20.

7. Improvements in or Relating to Magnetic Amplifiers, **C. S. Hudson.** British Patent *598,285*, February 13, 1948.

8. Saturating Core Devices (book), **L. R. Crow.** The Scientific Book Publishing Company, Vincennes, Indiana, 1949, pages 279–82 and 294–95.

9. Special Magnetic Alloys and Applications, **G. W. Elmen, E. A. Gaugler.** *Electrical Engineering* (New York, N. Y.), volume 67, September 1948, pages 843–45.

10. Magnetic Amplifiers with Orthonol Tape Cores, **W. A. Geyger.** *Proceedings*, National Electronics Conference (Chicago, Ill.), volume 6, 1950, pages 59–67.

11. Experimental Investigations on Magnetic Amplifiers for the Techniques of Measurement and Control, **W. Geyger.** *Wissenschaftliche Veroffentlichungen aus den Siemens-Werken* (Berlin, Germany), volume 20, 1942, figures 16 and 17, pages 38–41.

No Discussion

An Improved Magnetic Servo Amplifier

C. W. LUFCY
ASSOCIATE MEMBER AIEE

A. E. SCHMID
ASSOCIATE MEMBER AIEE

P. W. BARNHART
ASSOCIATE MEMBER AIEE

OF THE potential applications of magnetic amplifiers, one of the most promising is in the field of automatic feedback controls. In such systems, reliability of operation is frequently as important in design considerations as the operational characteristics themselves. For this reason, the replacement of relatively short-lived vacuum tube amplifiers with long-lived, rugged, and reliable magnetic amplifiers is an extremely desirable thing provided the minimum operational characteristics can be met using the magnetic amplifier.

In the past, one of the major difficulties in obtaining high performance from a magnetic amplifier servo system has been the slow speed of response of the amplifier itself. Although the magnetic amplifier is a carrier-type system and as such has a minimum response time of 1 cycle of the supply or carrier frequency, this minimum time is usually not obtained in conventional circuitry.[1] As a result, very serious stability problems are encountered if such an amplifier is placed in a high performance servo loop. Stabilizing or corrective networks, such as error rate, velocity error, acceleration, and so forth, as conventionally used in vacuum tube servo amplifier design to obtain improved operational performance and stability, cannot be used with any degree of success when directly applied to magnetic amplifier circuitry. This is due in large part to the inherent low impedance of the circuits involved.

Despite these drawbacks, the increase in reliability, ruggedness, and so forth, obtained by using magnetic amplifiers has made their use not only desirable but frequently necessary in many applications.

The magnetic amplifier and servo design group of the Magnetics Division of the Naval Ordnance Laboratory has been interested primarily in designing and developing magnetic amplifiers suitable for use in high-performance instrument-type servos for remote positioning applications. An instrument servo is defined as one in which the load inertia and friction, reflected through the gear train to the motor shaft, is negligible compared to inertia and friction of the motor itself. This report covers the development of a new type magnetic amplifier circuitry and gives design and performance data of typical servos of the aforementioned type using this new circuitry.

The Conventional Magnetic Servo Amplifier

Instrument servos of the type being considered generally use a 2-phase induction motor as the power drive and synchro units for error detection. It is thus necessary that the servo amplifier be capable of receiving a phase reversible a-c signal and delivering a phase reversible a-c output, the magnitude being proportional to the input magnitude and the phase being determined by the input phase. A simple single-stage magnetic servo amplifier using conventional circuitry is shown in Figure 1. This circuit which delivers a full-wave phase sensitive output is based on the parallel self-saturating circuitry of Logan.[2] Variations and modifications of this circuitry have in the past formed the basis of most magnetic servo amplifier designs.

This basic circuitry, however, has certain drawbacks for use in a servo amplifier. For the most part, these drawbacks constitute features which contribute to the over-all time constant or speed of response of the amplifier. The magnetic amplifier is a carrier-type amplifier and as such will have inherently a minimum response time of 1 cycle of the power or carrier frequency. However, using this conventional circuitry, this minimum response time is usually not achieved. Instead, a delay of from 2 to 20, or more, cycles is to be expected. A major factor is the L/R time constant of the control circuit. Referring to Figure 1, it is evident that the control is established through series-connected windings which encompass all four reactors.

The inductive impedance Z_L of any given control winding may be reduced greatly when the reactor upon which it is wound goes into saturation. However, at no time during the operating cycle will this occur on more than two of the control windings simultaneously. Therefore, the total Z_L of two of the windings will be the minimum ever presented to the over-all control circuit during a given operating cycle. The ratio L/R may be, and conventionally is, decreased by addition of a series resistance in the control circuit (R_C, Figure 1). In this manner, improved speed of response may be obtained. The inclusion of R_C, however, may decrease the over-all gain of the amplifier tremendously. Its inclusion also requires that control turns be the major factor in building up gain for the system since its presence tends to limit the control current. To increase control ampere turns with the current limited, requires additional control turns. It is thus not uncommon to find control windings running well into the thousands of turns in amplifiers of such design. Certainly the more control turns per reactor, the higher the control circuit Z_L. This in turn means a higher R_C if the L/R time constant is to remain unchanged. It is evident, therefore, that speed of response and gain are not independent variables.[3,4] Unfortunately, the gain requirements for most servo amplifiers are such that they cannot be met without seriously increasing the speed of response under these conditions.

In addition to the control circuit time constants, load and, if present, bias circuit time constants may be detrimental. These are, however, usually small compared to the control L/R and are generally neglected. Another possible source of delay is the presence of circulating currents within the amplifier itself. Referring again to Figure 1, it can be seen readily that internal d-c paths are present, for example, $A\ B\ C\ D$ in the power windings. Power consumed by circulating currents in these paths is derived in large part directly from mutual coupling with the control source. This power, of course, is not available for actual flux control within the reactors. Experience has definitely indicated that minimizing all induced direct currents in these paths, or elimination of them entirely, leads to a more efficient system.

Despite the aforementioned difficulties encountered when using the conventional magnetic amplifier circuitry, it should be pointed out that through careful design and engineering, servos can be built using such amplifiers which perform very satisfactorily within their design limits.[5]

The Reduction of Magnetic Amplifier Time Constant

In many cases, servo requirements are such that no excess time delay can be al-

---

Paper 52-235, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted March 28, 1952; made available for printing May 14, 1952.

C. W. Lufcy, A. E. Schmid, and P. W. Barnhart are with the United States Naval Ordnance Laboratory, Silver Spring, Md.

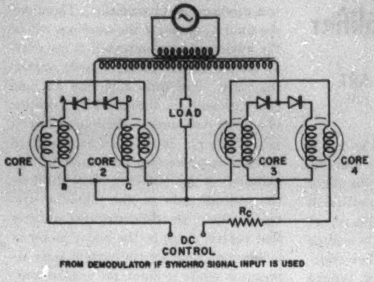


Figure 1 (left). A simple single-stage magnetic servo amplifier

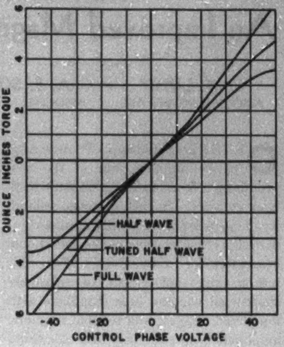


Figure 3 (right). Torque-voltage characteristics of 2-phase induction servo motor

lowed within the amplifier itself. One of the major problems in magnetic servo amplifier design and development, therefore, has been that of reducing the time constant, ideally, and if possible, to the absolute minimum of 1 cycle per stage. This must be done, of course, without undue sacrifice in gain.

The approach to this problem was to reduce in any way feasible the control winding L/R time constant. As pointed out previously, Z_L in any given control winding drops to a very low value when the reactor upon which it is wound goes into saturation. As long as conventional full-wave circuits are used, separation of the control windings such that only saturated, or unsaturated, cores are involved is difficult. If, on the other hand, only half-wave output is desired, the objective is achieved readily. Referring to Figure 1, if cores number 2 and number 4 are removed, the output is half-wave, but still with phase reversibility. Such a circuit is shown in Figure 2. In normal operation both cores will "fire," or saturate, during the operating half-cycle (that is, the half-cycle when voltage appears across the load). It can be seen readily that during this period the Z_L in both control windings is reduced to a very low value. Thus L/R drops to a low value and a definite control current is established. Actually, in the interval between the firing of one core and the other, the control winding on the unfired core acts as a voltage source due to mutual coupling with the power circuit. The voltage induced in the control circuit is always in such a direction as to aid control current build-up.

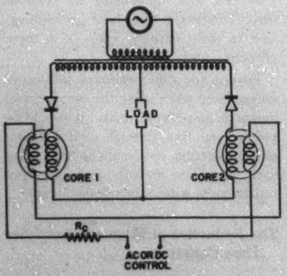


Figure 2. Half-wave magnetic servo amplifier with center-tapped power transformer

Since during the next operating half-cycle no output is obtained, this half-cycle is reserved for the purpose of controlling or presetting of flux. Thus the control current flowing at the end of the operating half-cycle is the current available at the beginning of the nonoperating or flux setting half-cycle. At any given level of operation, this same current is always established when both cores fire. If the same current is available at the beginning of each flux setting half-cycle, it is apparent that whatever flux level this current is capable of establishing during this period, will be established during each succeeding flux setting half-cycle. The operation is cyclical and, furthermore, no d-c transients in the control circuit are allowed to last longer than the period between the times when both cores are fired at the same time. It follows that the response of the system to a given control signal cannot exceed 1 cycle.

Following this same reasoning, it is apparent that a properly phased a-c control signal, or a properly phased pulse control signal will be effective in the half-wave circuit. It need only be established that at any given signal level the same voltage time integral is applied to the control winding during the nonoperating or flux setting half-cycle. If this is true, a cyclic operation is obtained in which no transients last longer than 1 cycle, thus assuring 1-cycle speed of response.

The fact that either direct current, alternating current or pulses, or for that matter combinations of these, are usable as control signals, makes the half-wave circuit quite flexible from the standpoint of control source requirements. Since at zero signal, no voltage appears across the output or load, the circuit is also compatible to cascading one stage on another. One drawback could be that the half-wave output would not be suitable for driving a 2-phase a-c induction motor. Although this output is unidirectional for any given phase sense input, the wave form which is applied to the motor will contain a high fundamental a-c component. Tests made on a typical 2-phase induction motor prove, however, that the torque obtainable with a half-wave signal is at least half that obtained with full-wave, and somewhat greater if capacitors are used across the motor to tune it to the fundamental frequency. Figure 3 shows the torque versus voltage curves of a 2-phase induction servo motor (5-watt low-inertia servo motor, Ford Instrument Company) for half-wave, tuned half-wave, and full-wave voltage.

The half-wave circuit fulfills the basic requirements for a servo amplifier, and in addition possesses the minimum speed of response possible for a magnetic amplifier. For these reasons this investigation of half-wave circuitry for use in high performance instrument-type servos has been carried out. The circuits developed and servo performance obtained using them are herein described.

Description of the Half-Wave Magnetic Servo Amplifier

THE BRIDGE CIRCUIT

The basic half-wave circuit shown in Figure 2 was obtained, as previously mentioned, by removing two reactors from the conventional full-wave parallel self-saturating circuit of Figure 1. This

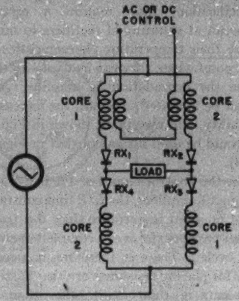


Figure 4. The basic half-wave magnetic servo amplifier bridge circuit

circuit possesses the advantages mentioned for use as a servo amplifier and, therefore, may be used as such. However, in actual design considerations it is desirable to keep components to an absolute minimum. One saving in this respect is to eliminate the center tapped power transformer by going to a bridge arrangement. Such a bridge circuit is shown in Figure 4. This circuit represents the basic design used in developing half-wave servo amplifiers. As indicated in Figure 4, opposite legs of the bridge are wound on the same reactor cores. Hence, only two cores are necessary although four could be used. It is preferable to wind the opposite bridge legs on the same core since for balanced operation they must act together. In this way, uniform operation is insured. For half-wave operation only two rectifiers, either RX_1 and RX_3, or RX_2 and RX_4 are absolutely necessary. In this circuit, however, all four rectifiers are used to completely eliminate any possibility for circulating currents in any leg or section of the bridge. It is to be noted that the load circuit, which may be either the control winding of the succeeding stage or the motor, is completely isolated from the bridge proper in that voltages induced in the load circuit cannot produce currents in the bridge.

Control of the Bridge Circuit

In the half-wave circuit the action of the control can be considered in a much simpler manner than in full-wave circuits. This is true since a complete half-cycle, the flux setting half-cycle, is available for the sole purpose of establishing control. During this period, the power winding circuit is inactive and its effect upon the control may be neglected. In full-wave circuits this, of course, cannot be done since some of the control windings are on reactors which are in their operating half-cycle.

Fundamentally, the requirement of the control is to establish a certain desired flux level in the reactor core during the flux setting half-cycle. This flux level essentially determines the firing angle of the reactor on its operating half-cycle.[8] Figure 5(A) shows a hysteresis loop which may be assumed to represent either reactor in the bridge because both are cycling around their respective loops in the same way at the same time if the bridge is in balance (that is, zero output). At the end of the operating half-cycle the reactor flux returns to residual point A. The control is then required to set a flux level, depending upon output desired somewhere between, say point D and C during the next half-cycle. It is not necessary that all the core losses involved in setting this flux level be supplied by the control source. In Figure 5(B) the basic bridge circuit is shown with the addition of another set of windings. These windings, called the reference windings serve the purpose of presetting a definite flux level, say B, in each reactor during the flux setting half-cycle. The power required to do this is obtained from the main a-c power source. The control source must then supply only the incremental power required to override the reference winding and shift the flux either up or down from point B. If the core material is quite rectangular, for example Orthonol, this incremental power may be very small. The fact that the control source is no longer required to overcome the coercive force of the core material opens the way for use of core materials having relatively large coercive force even in low level stages. The main requirement is that the material possesses a high incremental a-c permeability.

With reference to Figure 5(B) the action of the reference winding is as follows:

The main bridge power windings N_1, N_2, N_3, N_4 are active only when the a-c line polarity is positive at point X. (Determined by rectifiers RX_1, RX_2, RX_3, RX_4.) At the beginning of the operating half-cycle, that is, when point X goes positive, the core flux is at, say point B, Figure 5(A). The flux is then carried up the hysteresis loop to saturation where it remains until the end of this half-cycle, at which time the flux returns to residual, point A. During this period, no current flows in the reference windings N_{r1}, N_{r2} because of rectifiers RX_5, RX_6. The next half-cycle, however, line polarity is reversed and no current flows in windings N_1, N_2, N_3, N_4 but does in windings N_{r1}, N_{r2}. The currents flowing through N_{r1} and N_{r2} set up a magnetomotive force on the reactors opposite to that in the main power windings. Hence, the flux change brought about in the reactor is opposite. Thus, with zero control winding current the flux can be made to move to a predetermined level by the action of the reference windings. The required flux reference level, say point B, is set independently on each reactor by proper adjustment of the number of turns on the reference windings and the series resistance R_1 and R_2. In practice, rectifiers RX_5 and RX_6 may be eliminated if N_{r1}, N_{r2} and R_1, R_2 are properly determined. However, control of exact reference flux level is more critical without them.

The fact that each reactor may be independently adjusted or preset to the same reference flux level, regardless of the core's magnetic properties, is of considerable advantage in obtaining balanced operation from cores that normally are too unlike, magnetically, to be used in the same circuit. The "core matching" problem of course, is not completely eliminated. If two cores are dissimilar, although the same reference flux level may be readily established, the incremental flux change brought about by the control winding itself may be quite different. The important thing is that the incremental a-c permeability of the two cores

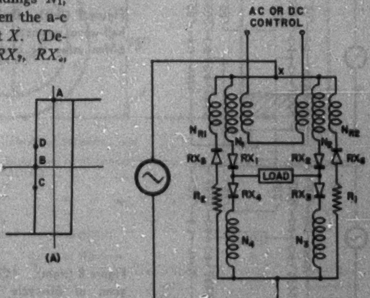


Figure 5(A). Hysteresis loop of reactor cores. (B). The half-wave bridge circuit with reference winding

be approximately matched in a region about the reference flux level so that the control has an equal effect on each reactor. The coercive force of the cores may be dissimilar and the actual shape of the a-c hysteresis loop near the saturation level may be dissimilar and balanced operation still obtained by proper adjustment of the reference windings

It would appear from the foregoing discussion that a separate reference winding is not necessary. A controlled back magnetomotive force, as furnished by the reference winding, could be obtained by shunting rectifiers RX_1, RX_2, RX_3, RX_4 and allowing a reverse current to flow through the power windings N_1, N_2, N_3, N_4. (A certain amount of this current does flow due to rectifier back leakage in actual practice. In the discussion, however, this has been neglected to simplify the argument.) The circuit would then be as shown in Figure 6. Actual measurement on these two circuits, however, shows a reduction in gain when shunted rectifiers are used. This is believed to be due to circulating currents which can flow around the bridge through these shunts. The normal turns ratio between the control windings and the power windings is such that the reflected impedance of these shunts is such as to make them appear as a noticeable load to the control circuit. On the other hand, the turns ratio between the control and reference windings can be made such that the impedance in the reference circuit is reflected as a high impedance into the control circuit. It is pointed out that back leakage in the power circuit rectifiers, whether due to deliberate shunting or through inefficiency in the rectifier, cause a flux change in the desired direction, that is, down the back side of the hysteresis loop. It is not the serious matter it can be in the conventional full-wave self-saturating circuit where the degree of positive feedback is determined by rectifier leakage. Frequently, ambient temperature variations, required in a given specification, may require a certain amount of shunting of rectifiers to minimize their temperature characteristics to the point where they can be used. In this case the degradation in amplifier gain must be tolerated to obtain the increased stability, and the circuit shown in Figure 6 would be used instead of that in Figure 5(B).

In the control circuit of the half-wave magnetic amplifier, the L/R time constant is of no great importance since d-c transients cannot exist over a complete operating cycle.[8] There is, therefore, no necessity for placing resistance or other passive elements in this circuit other than as might be demanded by the control source itself. Where synchro components are used for error detection in the servo system, it is frequently desirable to keep a minimum impedance of 10,000 ohms in the input circuit. It is conventional

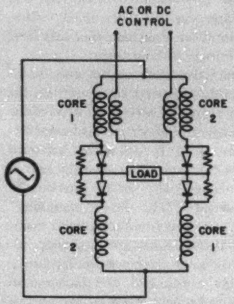


Figure 6 (left). The half-wave bridge circuit with shunted rectifiers

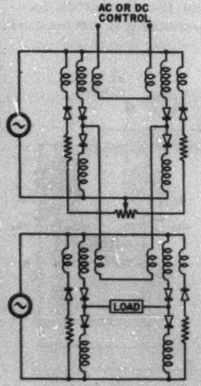


Figure 7 (left). Basic 2-stage half-wave magnetic servo amplifier using the bridge circuit

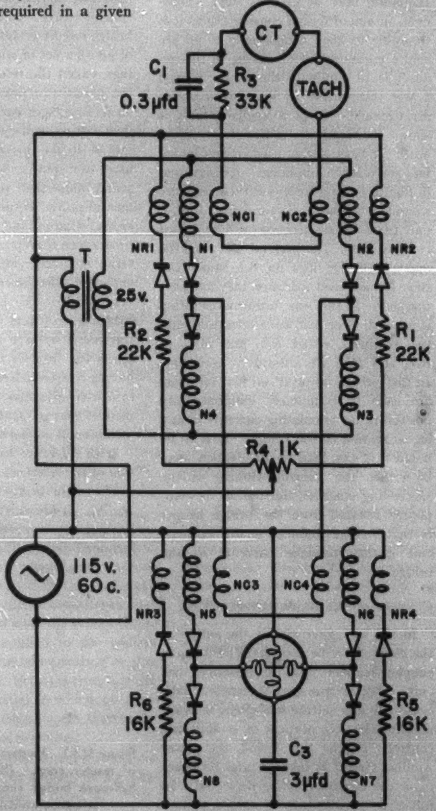


Figure 8 (right). Circuit diagram of 60-cycle half-wave magnetic servo amplifier

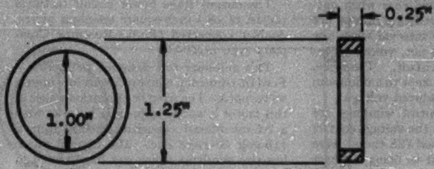


Figure 9. Core and winding data for 60-cycle amplifier input stage

Details of Magnetic Cores

Material	Orthonol (50-per-cent Ni 50-per-cent Fe alloy, grain oriented, dry hydrogen anneal)
Configuration	Tape wound toroid
Inside diameter	1.00 inch
Outside diameter	1.25 inches
Tape, width	0.25 inch
Tape, thickness	0.002 inch
Insulation	MgO applied in winding process

Details of Windings

Windings	Function	Number Turns	Wire Size
N_1 N_2 N_3N_4	Power	2,000	Number 33
N_{R1} N_{R2}	Reference	500	Number 36
N_{C1} N_{C2}	Control	100	Number 36

Each core has two power windings, a reference, and a control winding

practice to insure this minimum by insertion of a resistance in the input control circuit. The disadvantage of doing this is loss of gain and limiting of current in the control winding, as has been previously discussed. A large number of turns on the control windings is usually required to obtain a reasonable gain because ampere turns of control becomes the determining factor.

Increased Gain Through Reduction in Control Turns

To set down equations for flux level in terms of ampere turns magnetomotive force on a reactor having a highly rectangular hysteresis loop is virtually impossible, since the ampere turn, or H, and flux level B relationship is nearly discontinuous in many magnetic materials available today. This difficulty has been pointed out clearly and a more satisfactory analysis of magnetic amplifier performance in terms of voltage time integrals has been made by R. A. Ramey.[7] In order to make use of this method, however, it is necessary that the voltage time integral which appears on the windings of the reactors be known. Normally, this information is as difficult to obtain precisely as is the flux level in terms of ampere turns. The instantaneous currents flowing in a circuit involving resistance and reactance due to a magnetic amplifier reactor are so complex that although the voltage time integral applied to the system externally may be known exactly, the portion of this which finally appears across the reactor winding itself, and which uniquely determines $d\phi/dt$ in the core, is difficult to establish. The basic approach, or method of analysis is, nevertheless, extremely useful in terms of understanding what actually takes place in the magnetic amplifier.

Ideally, if a control circuit contained no resistance and the control voltage time integral were known, the amount of flux change which would be obtained by this control is given by Faraday's law as

$$\Delta\phi = \frac{1}{N}\int E_c dt$$

where

N = Number of turns of control winding
$\int E_c dt$ = Voltage time integral over the flux setting half-cycle (for the half-wave circuit)

For a given level of control, that is $\int E_c dt$ fixed, the maximum flux change would be obtained by making N as small as possible. In other words, if voltage control is utilized, the voltage gain of an amplifier is increased if the number of control turns is decreased. To insure voltage control, it is necessary that the control be a low impedance source. Of course, in practice it is possible to get so low in control turns that the current drain on the control source necessary to maintain a given voltage across the control windings may become excessive. There will be, consequently, an optimum number of turns on the control winding representing a compromise between the voltage and the current limitations of a given source.

The utilization of voltage control in the input control circuit, where resistance is usually present due to control source limitations, is not possible. However, the 2-stage bridge circuit lends itself to this principle very nicely. The output of the first stage may be fed directly into the control winding of the second stage with no passive element required. Thus, one may be assured that whatever voltage appears at the output of the first stage is impressed directly into the control winding of the second.

Conclusion

A 2-stage magnetic servo amplifier has been developed utilizing the principles set forth in the foregoing discussion. The basic circuit is shown in Figure 7. This circuit has been used in both 60- and 400-cycle applications to obtain performance much superior to that obtainable with conventional full-wave circuitry. The over-all maximum response time of this circuit is $1^1/_2$ cycles of the power frequency (1 cycle for first stage, 1/2 cycle

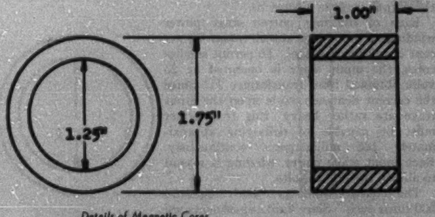


Figure 10. Core and winding data for 60-cycle amplifier output stage

Details of Magnetic Cores

Material	Orthonol (50-per-cent Ni 50-per-cent Fe alloy, grain oriented, dry hydrogen anneal)
Configuration	Tape wound toroid
Inside diameter	1.25 inches
Outside diameter	1.75 inches
Tape, width	1.00 inch
Tape, thickness	0.002 inch
Insulation	MgO applied in winding process

Details of Windings

Windings	Function	Number Turns	Wire Size
N_5 N_6 N_7 N_8	Power	1,500	Number 30
N_{R3} N_{R4}	Reference	500	Number 34
N_{C3} N_{C4}	Control	25	Number 28

Each core has two power windings, a reference, and a control winding for each additional stage). The ease with which additional stages may be cascaded is, of course, evident. In comparison to conventional full-wave circuitry only about one-half as many components, occupying approximately one-half the volume, are required.

The complete design and adjustment details for typical 60- and 400-cycle half-wave servo amplifiers are shown in Appendix I, and the performance characteristics of the two complete servo systems are shown in Appendix II.

This report has been written purposely in a general nature to introduce the basic concepts underlying the development of the amplifier circuitry described. The utilization of half-wave magnetic amplifiers for instrument servos represents a new approach to this problem. The fact that delay time in the amplifier has been reduced to the absolute minimum opens the way for further development and utilization of magnetic amplifiers in high performance servo systems heretofore requiring vacuum tube amplifiers.

Appendix I. Design and Adjustment Details for Typical 60- and 400-Cycle Half-Wave Servo Amplifier

60-Cycle Half-Wave Servo Amplifier Circuit

The complete circuit diagram of a 60-cycle servo amplifier based on the half-wave bridge is shown in Figure 8. This circuit will be recognized as basically that of Figure 7. All cores are of a 50-per-cent iron 50-per-cent nickel grain oriented material having a very rectangular hysteresis loop. The exact core and winding data for the input stage are given in Figure 9 and for the output stage in Figure 10.

Each of the four output stage power windings is wound to absorb about 75 per cent of full-line voltage. To permit smaller cores, the input stage is operated at 25 volts obtained from transformer *T*. Since the current demands made upon the input stage are rather heavy, this transformer must be capable of delivering approximately 125 milliamperes continuously. Each input stage power winding is wound to absorb the full 25 volts.

The reference windings of both stages are 500 turns each. Since final adjustments are made in the series resistors in the reference circuits, this number is not critical. It should be high enough that excessive currents are not required for referencing, but not so high that valuable winding area is sacrificed. The mutual coupling between the reference winding and control circuit becomes critical also if reference turns are too great.

The input stage control windings are 100 turns each; however, a higher gain would result in the amplifier if these turns were increased. It is recognized that when output is desired, the two reactors do not fire together and during the interval when one is fired and the other is not, voltage is induced in the control circuit. Therefore, these windings should be kept to a minimum to reduce this value of induced voltage.

The output stage control windings are very low in keeping with the voltage control concept. The values used (25 turns) were experimentally arrived at as being the best compromise for the input stage shown. However, the operation of the amplifier is not noticeably changed if this number is reduced to 8 or 10 turns.

All fixed resistors shown are 1 watt carbon, each experimentally determined for the particular core and its desired operating conditions. R_4 is a 4,000-ohm potentiometer used to bring the amplifier into balance.

The network R_3C_1 in the input circuit is used primarily to guarantee a minimum of 10,000 ohms input impedance. The *RC* combination allows a slight phase adjustment found to be helpful in obtaining optimum operation, which a simple resistance would not permit.

The rectifiers in the power circuit of the input stage consist of four 5/8-inch-square selenium plates, doubler connected (Radio Receptor Company type *J* plates in intimate contact). The input and output stage reference winding rectifiers consist of eight 1/2-inch-square selenium plates, half-wave connected (Radio Receptor Company type *Y* plates in intimate contact).

The output stage power circuit rectifiers consist of 18 1-inch-square selenium plates, doubler connected (Radio Receptor Company type *18M0488*).

This amplifier was designed to operate a Ford Instrument Company 5-watt low-inertia servomotor. The fixed or excitation phase of this motor is wound to operate in series with a 3.5-microfarad capacitor directly from a 115-volt 60-cycle line. The control phase has two windings which can be operated in series at 115 volts or in parallel at 57.5 volts. In this design, the control windings are used in parallel presenting an impedance of about 250 ohms to the amplifier output.

The amplifier is adjusted to operate about a quiescent firing angle (that is, point during operating half-cycle where cores saturated with zero signal input) of 90 degrees in the input stage and approximately 150 degrees in the output stage. These quiescent firing angles are set individually on each core by adjusting the series resistors in the reference windings. It is preferable to operate about a 90 degree

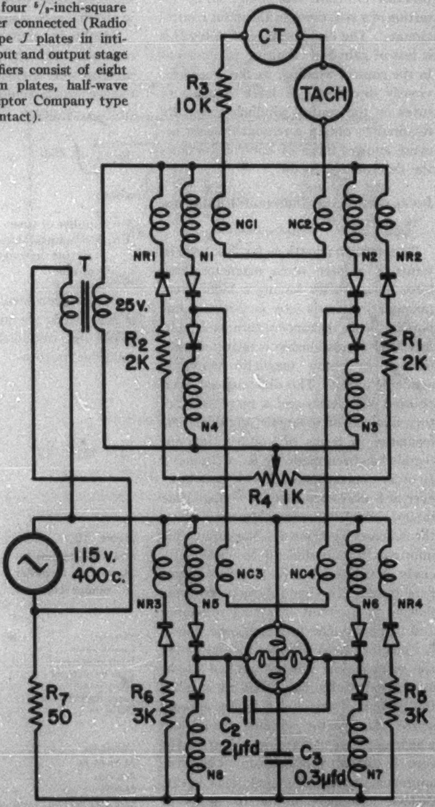


Figure 11. Circuit diagram of 400-cycle magnetic servo amplifier

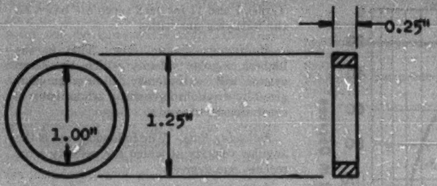


Figure 12. Core and winding data for 400-cycle amplifier input stage

Details of Magnetic Cores

Material	Orthonol (50-per-cent Ni 50-per-cent Fe alloy, grain oriented, dry hydrogen anneal)
Configuration	Tape wound toroid
Inside diameter	1.00 inch
Outside diameter	1.25 inches
Tape, width	0.25 inch
Tape, thickness	0.002 inch
Insulation	MgO applied in winding process

Details of Windings

Windings	Function	Number Turns	Wire Size
N_1 N_2 N_3 N_4	Power	500	Number 28
N_{R1} N_{R2}	Reference	50	Number 31
N_{C1} N_{C2}	Control	300	Number 36

Each core has two power windings, a reference, and a control winding

firing angle in each stage, since at this point maximum output is obtained for a given differential in the firing angle of each core. It is not desirable to operate the output stage of this amplifier about 90 degrees, however, because the quiescent currents flowing around the bridge are excessively high causing heating of the rectifiers and power winding. For this reason, the output stage quiescent current is reduced to a lower value by firing the cores at about 150 degrees. The over-all efficiency of any given stage can be increased by operating the stage closer to cutoff value (that is, firing angle of 180 degrees). If the stage is never to be operated at cutoff, it is not necessary that each power winding absorb full-line voltage but only the voltage time integral necessary to maintain the desired firing angle. The maximum output cannot be obtained, however, unless each winding will absorb full-line voltage. In this connection, it should be pointed out that the operation of a given stage is never symmetrical. If the firing angle of one core is advanced, say 5 degrees, by the control action, the other core is not retarded 5 degrees. This is true because when the first core fires full-line voltage, less drop across series rectifier and winding of first core, appears across each winding of the second. On the other hand, before the first core fires line voltage appears across both windings of the second in series. Hence, when one core fires, $d\phi/dt$ in the other is practically doubled so that it fires sooner than it would normally if the line voltage always appeared in series with its power windings instead of suddenly in parallel with them.

The choice of core size and power winding turns is determined primarily on the basis of line voltage to be controlled and the load current requirements. For example, if full-line voltage is to be controlled, the core size and power winding turns must be such as to absorb approximately 115 volts on each winding.

This can be accomplished by any combination of core size and power winding turns which will satisfy the following well-known expression

$$E=\frac{4.44BNAf}{10^8}$$

where

E = voltage absorbed by reactor
B = maximum flux density
N = number of turns on power winding
A = core cross-sectional area
f = power frequency

However, in designing a reactor, the power winding size and the number of turns are held to values which will produce a low-voltage drop across the reactor during saturation.

The procedure is usually, first, to determine the power winding wire size necessary to carry the required load current without excessive heat losses, then, to use a core size which has a sufficient window area to allow winding the required number of turns of this wire size. The number of turns, of course, depends upon the core size, but normally only one or two test calculations are necessary before the right combination of core dimensions and winding turns are arrived at. Where standard core sizes are available, the nearest core size to meet these requirements would be chosen without concern over whether this particular core size to copper area ratio is the exact optimum.

400-Cycle Half-Wave Servo Amplifier Circuit

The complete circuit diagram of a half-wave 400-cycle amplifier is shown in Figure 11. Again the similarity to Figure 7, and Figure 8, is evident. The general discussion in the previous section relating to design and adjustment considerations of the 60-cycle amplifier apply also to this circuit, hence, will not be reiterated. The exact core and winding data for the 400-cycle input stage are shown in Figure 12, and for the output stage in Figure 13. Transformer T must deliver 25 volts at 250 milliamperes. All resistors, with the exception of R_p, are 1 watt carbon, each experimentally determined for the particular core and its desired operating conditions. R_p is a 50-ohm 10-watt wire wound resistor. No input circuit RC network is used in the 400-cycle amplifier since phase shift adjustments in this unit were not found necessary.

In this amplifier, both input and output stage are adjusted to a quiescent firing angle of 90 degrees, for maximum performance. As previously stated, adjustment of the output stage to this quiescent level allows rather heavy currents to flow around the bridge. To reduce this current to a satisfactorily low value, without deviating from the 90-degree firing point, as was done in the 60-cycle amplifier, the series resistor R_p is inserted in the main power circuit. Although voltage dropped in this resistor is not available for motor operation, the loss here is more than offset in increased gain by operating the output stage at a 90-degree firing angle. Of course, R_p

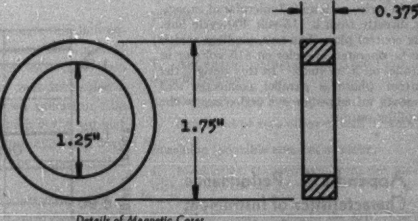


Figure 13. Core and winding data for 400-cycle amplifier output stage

Details of Magnetic Cores

Material	Orthonol (50-per-cent Ni 50-per-cent Fe alloy, grain oriented, dry hydrogen anneal)
Configuration	Tape wound toroid
Inside diameter	1.25 inches
Outside diameter	1.75 inches
Tape, width	0.375 inch
Tape, thickness	0.002 inch
Insulation	MgO applied in winding process

Details of Windings

Windings	Function	Number Turns	Wire Size
N_5 N_6 N_7 N_8	Power	850	Number 29
N_{R3} N_{R4}	Reference	50	Number 31
N_{C3} N_{C4}	Control	10	Number 22

Each core has two power windings, a reference, and a control winding

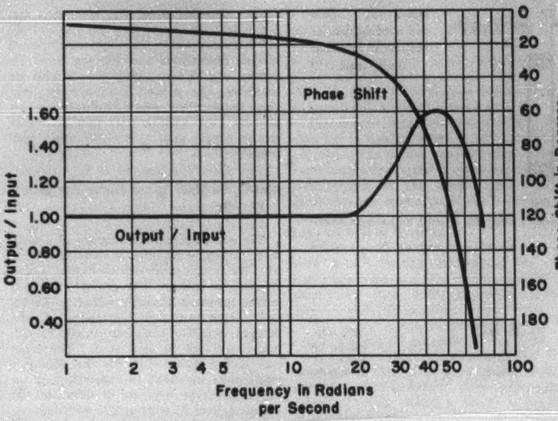


Figure 14. Closed-loop system response (60 cycle)

must be less than a certain maximum resistance if this is to continue to be true, for eventually the voltage drop across it would cause a decrease in over-all gain. The value for R_p shown, 50 ohms, was chosen to give maximum voltage gain of the amplifier.

The rectifiers in the power circuit of the input stage consist of two, 1-inch-square selenium plates, doubler connected (Radio Receptor Company type *2MO454*), and the input stage reference winding rectifiers are *1N34* germanium diodes.

The output stage power rectifiers consist of 12, 1-inch-square selenium plates, doubler connected (Radio Receptor Company type *12MO447*), and the output stage reference winding rectifiers are the same as in the 60-cycle amplifier reference circuits.

This amplifier was designed to operate a *Mk 16 Mod O*[8] low-inertia servo motor. The fixed phase of the Mk 16 motor operates in series with an 0.3 microfarad capacitor directly from a 115-volt 400-cycle line. The control phase has two windings which can be operated in series on 115 volts or in parallel on 5705 volts. In this design, the control phase is parallel connected and presents an impedance of 365 ohms to the amplifier output.

Appendix II. Performance Characteristics of Instrument Servos Using 60-Cycle and 400-Cycle Half-Wave Amplifiers

The 60-cycle amplifier described was evaluated in a closed-loop instrument-type servo of the remote positioning type. The gear reduction from the motor to the synchro current transformer shaft was 22.5-to-1 and the system was stabilized with an a-c tachometer generator (induction generator type *945B-01600*, Kollsman Instrument Corporation) geared directly to the motor. The tach output was fed back directly in series with the current transformer error signal.

The 400-cycle amplifier was evaluated in a similar closed loop servo. The gear reduction from the motor to the synchro current transformer shaft was 30-to-1. The system was stabilized with an a-c tachometer generator built directly on the motor shaft. This tachometer output was also fed directly in series with the current transformer error signal. (Advantage may be taken of the internal impedance of the tach so that if 10,000-ohm impedance is necessary in the current transformer rotor circuit the additional circuit impedance may be only that required to equal 10,000 ohms when added to the tach impedance.)

The performance characteristics given in Tables I and II for each servo are based on the following quantities

Maximum Slewing Rate. This is the highest angular velocity which the servo system will continuously follow when energized by a synchro generator whose rotor is continuously rotated.

Following Rate. The following rate is the angular velocity at which the velocity error reaches one degree.

Static Error. The static error is the maximum difference in angular position between the current transformer in the servo system and the synchro generator which excites it when the system is at rest. This figure does not include drift error.

Frequency Response. The frequency response data was obtained using servo frequency analysis equipment similar to that described in *NOLM 10930*.[9] A suppressed carrier signal of variable amplitude and frequency is applied to the system. The amplitude and phase of the output is then measured. From this data, system stability and performance are obtained by conventional methods.

Voltage Gain. This is the voltage gain of the magnetic amplifier with the current transformer utilized as a signal source and the servo motor (stalled torque conditions) as a load on the amplifier. Both motor and current-transformer voltages were read on a Ballantine model *300 VTVM*. This must be taken into consideration since the output wave form is nonsinusoidal and the indicated motor voltage depends upon the type meter used. The value shown is for small signal inputs within the linear range of operation of the amplifier.

All were taken at room temperature with line voltage of 115 volts 60 cycle (or 400 cycle). The gear trains were carefully adjusted to minimize back lash and sticking. Performance of the systems under extremes of voltage, frequency, and temperature as are required in military specifications are not herein given. The major difficulty encountered in meeting such specifications is almost always associated with drift in a

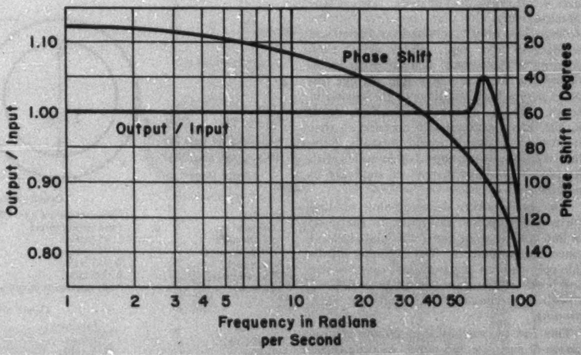


Figure 15. Closed-loop system response (400 cycle)

magnetic amplifier system. The performance of the half-wave amplifier in this respect will be comparable with previously reported tests for full-wave systems[6] because both are balanced bridge amplifiers and the causes of drift (bridge unbalance) are common to both. The elimination of drift is closely associated with core and rectifier matching. Of the two, the rectifiers are much more difficult to control and maintain. The data in Tables I and II are intended to give an idea of the dynamic servo performance characteristics of the half-wave magnetic amplifiers and should not be taken as the ultimate performance which might be obtained in a system completely engineered to meet rigorous military specifications.

Table I. Measured Performance of the 60-Cycle Servo

Maximum slewing rate	550 degrees per second
Following rate	1 degree per 60 degrees per second
Static error	less than 0.1 degree
Frequency response and phase shift	Figure 14
Voltage gain	12

Table II. Measured Performance of the 400-Cycle Servo

Maximum slewing rate	1000 degrees per second
Following rate	1 degree per 200 degrees per second
Static error	less than 0.01 degree
Frequency response and phase shift	Figure 15
Voltage gain	20

References

1. A STUDY OF DELAYS IN RESPONSE OF MAGNETIC AMPLIFIERS AS AFFECTED BY DIFFERENT CORE MATERIALS, POWER AMPLIFICATION FACTOR POWER FREQUENCIES, AND CIRCUITS, E. V. Weir. *Memorandum 10555*, United States Naval Ordnance Laboratory (Washington, D. C.), 1949.

2. ELECTRONIC CONTROLLING APPARATUS, F. G. Logan. United States Patent Number *2,126,799*, August 16, 1938.

3. APPLYING MAGNETIC AMPLIFIERS, PART III, W. J. Dornhoefer, V. H. Krummenacher. *Electrical Manufacturing* (New York, N. Y.), August, 1951.

4. SOME ASPECTS OF MAGNETIC AMPLIFIER TECHNIQUES, F. E. Butcher, R. Willheim. *Proceedings*, Institute of Radio Engineers (New York, N. Y.), March 1952.

5. THE APPLICATION OF A MAGNETIC AMPLIFIER TO A 60-CYCLE INDICATING SERVO SYSTEM, P. W. Barnhart. *NavOrd Report 1892*, United States Naval Ordnance Laboratory (Washington, D. C.), 1951.

6. THE APPLICATION OF A 60-CYCLE MAGNETIC AMPLIFIER TO A POSITION INDICATING SERVOMECHANISM, A. E. Schmid. *Proceedings*, National Electronics Conference (Chicago, Ill.), 1951.

7. ON THE MECHANICS OF MAGNETIC AMPLIFIER OPERATION, R. A. Ramey. *AIEE Transactions*, volume 70, part II, 1951, pages 1214–23.

8. CATALOGUE OF PRECISION INSTRUMENT COMPONENTS. *Publication OP1755*, United States Bureau of Ordnance (Washington, D. C.), section 10.12.

9. AN IMPROVED SERVOMECHANISM FREQUENCY RESPONSE ANALYZER, G. B. McCarter. *Memorandum 10930*, United States Naval Ordnance Laboratory (Washington, D. C.), May 17, 1950.

No Discussion

Electronic Recorder with Range and Precision Adequate for the Platinum Resistance Thermometer

ALBERT J. WILLIAMS, JR.
MEMBER AIEE

THE platinum resistance thermometer is the interpolating device for the 1948 International Temperature Scale from −182.97 degrees centigrade to +630.5 degrees centigrade.[1–3] When properly made the reproducibility of its resistance justifies its calibration and use for measurements to 0.01 degrees centigrade or 1/80,000th of its temperature range.[3] Mueller attacked this difficult resistance-measurement problem with a hand-operated decade bridge.[4] Stull attacked the problem with a mechanically balanced recorder using a slide-wire plus decades[5] and explained the value of the recorder, one example being the determination of chemical purity by freezing point measurement.[6] While Stull was building his recorder the author and his associates were also building a recorder, using one of the several methods of electronic balancing which had then been developed.[7,8] Based on the experience with these two types of recorders there was developed a third type with improved precision, the experimental form of which is the subject of this paper. The commercial form may differ in details from the experimental form described here.

The recorder has a slide-wire with a span of 1.02 ohms. A 1-ohm decade and a 10-ohm decade are automatically switched to give a total range to the recorder of 0 to 100 ohms. Features will be described which permit unattended operation of the recorder with a precision approaching 0.001 ohm, which corresponds approximately to 0.01 degree centigrade for a 25-ohm platinum resistance thermometer.

Unattended Operation and Precision

In the following text are listed three items which are required for unattended operation that would not be required if an operator were in attendance to do certain things. There are listed also three items affecting precision which would exist even if an operator were in attendance. Taken together, these six items form a check list found useful to the author, which may assist the reader in understanding the paper.

Unattended operation

1. Automatic changing and recording of decade resistors.
2. Compensation for expansion of chart.
3. Smallness of zero offset of null detector.

Precision (possible sources of error)

4. Mechanical transmission from the slide-wire contact to mark on chart.

Paper 52-157, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted January 11, 1952; made available for printing April 11, 1952.

ALBERT J. WILLIAMS, JR., is with the Leeds & Northrup Company, Philadelphia, Pa.

The author wishes to give credit to D. R. Stull who, with his associates at the Dow Chemical Company, pointed out the need for an instrument of this general type,[6] and to acknowledge the many significant contributions of the late Frank F. Dern, Jr.

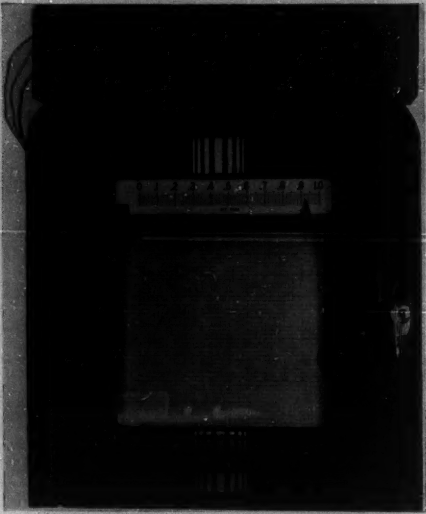


Figure 1 (left). Recorder and resistance box

Five figures may be read from the 5-dial box and five figures may be read from the recorder chart

Figure 3 (right). Printer and decade switches

The interconnection is the diagonal shaft at the top

5. Setting of slide-wire contact on true balance position.

6. Changes in bridge zero, in slide-wire or in decade resistors between calibration and use.

Record of Changes in 5-Decade External Resistance Box

Figure 1 shows the recorder and a resistance box. The resistance box is not a part of the recorder but was used in place of the platinum resistance thermometer for the following tests. Five figures may be read from the 5-dial resistance box and five figures may be read from the recorder chart.

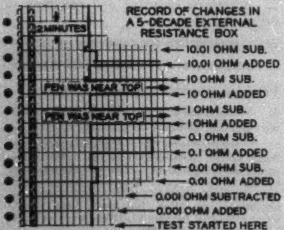


Figure 2. Record of changes in 5-decade external resistance box

Small values of resistance may be read from the pen line. Large values of resistance require reading of the printed figures also

The record of changes made in the 5-decade external resistance box, as shown in Figure 2, seems to require no explanation for the addition and subtraction of the 0.001-ohm, 0.01-ohm, and 0.1-ohm steps from the external resistance box. For the 1-ohm step, the pen moved 100 divisions and stopped to record at or very near to the 1.00-ohm chart line. For the 10-ohm step, the pen also stopped to record at or very near to the 1.00-ohm chart line, the other 9 ohms being recorded by the right-hand or units print wheel. Subtraction of the 10 ohms returned the print wheel and then the pen to zero. The print-wheel change was initiated by travel of the pen, and slide-wire shaft, to a position below zero, as shown on the chart. When 10.01 ohms were added, the pen moved to its uppermost position and remained there until the decades moved to 1 in the tens column and 0 in the units column, thus recording 10 ohms. The 0.01 ohm was recorded by the pen, coming in from the high side. Subtraction of the 10.01 ohms from the external box returned the print wheels and pen to zero.

Printer and Decade Switches

For Figure 3, the chart and chart roll have been pulled away to show the two print wheels in the center of the picture. The 1-ohm and 10-ohm decade switches are at the right. The mechanical interconnection between the switches and the print-wheel unit is the diagonal shaft at the top. The inking roller is just visible in its retracted position behind the print wheels.

A reversible capacitor-type motor drives the print wheels through gears and thence through a shaft shown at the top of Figure 4. The Geneva movement at the right drives the 10-ohm decade on the bottom shaft. The star wheel on the 1-ohm shaft operates an "index" switch which assists in stopping the motor. The trip switches, at the extreme lower left, start the motor in the proper direction whenever the pen approaches the upper or the lower limit of its travel.

The trip switches are shown from the rear in Figure 5. The arms on the slide-wire shaft are set to allow the pen to travel a little more than 1.00 ohm without tripping, to avoid frequent changes of the decades when the thermometer resistance is hovering around some integer value.

TRIPPING SWITCHES

Figure 4. Drive for print wheels and decade switches

The gears drive the print wheels through shaft at top. The Geneva movement, at right, drives 10-ohm decade on bottom shaft. The star wheel on the 1-ohm shaft operates the index switch

Figure 5 (left). Trip switches and actuating arms on slide-wire shaft

The arms are set to allow a slight amount of overtravel at both ends of slide-wire

Figure 6 (right). Decade-motor circuit

The index switch gives push-button action, braking, and precise positioning. The friction switch remembers the direction of motor start and so avoids the use of relays

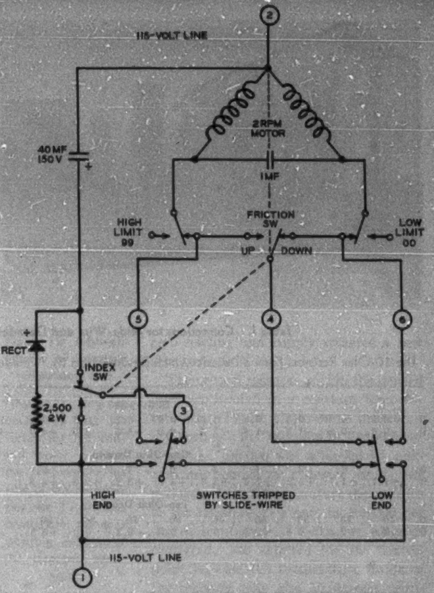


Figure 6 shows the decade-motor circuit. The index switch brakes the motor with direct current when the decade switches and print wheels have moved into the desired position. The "friction" switch reverses with reversal of the motor thus remembers the direction of motor start and so avoids the use of relays.

Humidity Compensator

Strip charts made of paper expand in width by several tenths of a per cent with increase of humidity and so could introduce errors of several tenths of a per cent. To avoid such errors, mechanical means as shown on Figure 7 have been used so that expansion of the chart moves the pen as well as the chart thus preventing relative motion from this cause.

Harrison[9] had this basic idea over 20 years ago. The execution of the idea for this recorder is different[10] and is most readily presented by Figure 8. The spider is positioned axially by the chart. Through the groove in its hub, via the engaging pin, a bell crank and a long lever running through the pen carriage to a pivot near to the other edge of the chart, the pen is moved in space with expansion of the chart so as to avoid motion relative to the chart. The long lever is pivoted near to the fixed edge of the chart so that its effect on the pen will be proportional to the pen's distance from the fixed edge of the chart, as it should be to avoid relative motion. The chart at right of Figure 8 shows the performance with increasing, then decreasing humidity. The line made by a fixed pen shows that the chart expanded about one-half of a division or 0.5 per cent, but the mark made by the compensated pen shows no change.

The Balancing Amplifier

The balancing amplifier as shown in Figure 9 has an extra stage of amplification which provides sufficient extra sensitivity to allow reduction of the thermometer current, from the 5 to 8 milliamperes which Stull[3] was forced to use, to 2 milliamperes as used by the National Bureau of Standards for calibration of precise platinum thermometers suitable for this recorder. Some of the spare gain can be used to make the motor responsive to random noise thus eliminating any persisting effect from friction and so obtaining extreme sensitivity useful for self-checking as will be explained later.

Any spurious voltage or noise which when referred to the input has a d-c equivalent or low-frequency equivalent is

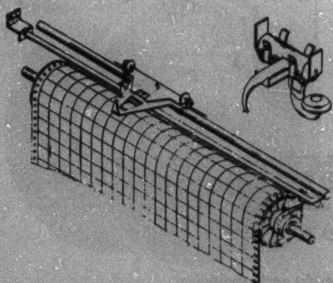
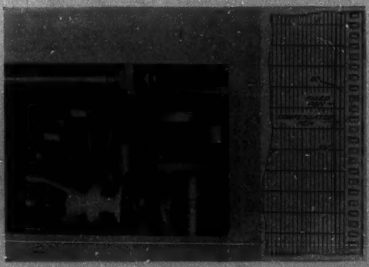


Figure 7 (left). Humidity compensator. Expansion of the chart moves the pen, thus preventing relative motion

Figure 8 (right). Humidity compensator. To the left is shown actuating spider positioned axially by the chart. The chart shows performance with increase, then decrease, of humidity



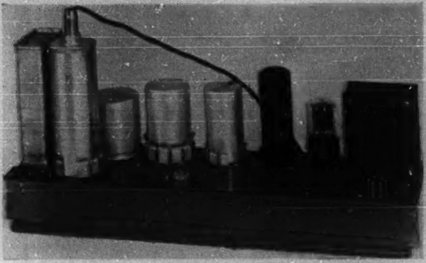


Figure 9. Balancing amplifier

Energization of the vibrator through its top assists in reducing zero offset to a negligible value

objectionable because it does not average out, hence is a possible source of error. Such so-called "zero offset" of "zero error" in this amplifier is negligible with respect to the 2 microvolts which corresponds to 1/10 division on the chart or 0.001 ohm.

The circuit of the balancing amplifier is shown in Figure 10. The five stages of amplification are contained in three envelopes by the use of two twin triodes.

Many things had to be done to reduce the zero offset to a negligible value. One of them, evident from this circuit, is the use of direct current on the heaters. Many details in the amplifier are the same as those described in an earlier paper.[11]

The resistance-capacitance network at the left provides filtering to exclude a-c noise and provides damping to combat the tendency of the balancing system to overshoot the balance point.

Bridge Circuit

Figure 11 shows the bridge circuit. Connected to the slide-wire at the left is the balancing amplifier consisting of the

Table I. Corrections for Slide-Wire and Decades

The 10-Ohm Resistors Form a Decade which Deviates from its Average by only 0.0006 Ohm

Slide-wire											
0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	reading in ohms
0	−0.6	−0.8	−1.3	−1.8	−2.8	−3.0	−3.0	−3.8	−4.3	−4.3	correction in milliohms
One-Ohm Decade											
0	1	2	3	4	5	6	7	8	9		reading in ohms
0	+1.1	+1.7	+2.2	+2.7	+3.2	+3.2	+3.2	+2.3	+2.8		correction in milliohms
Ten-Ohm Decade											
0	10	20	30	40	50	60	70	80	90		reading in ohms
0	+0.6	+0.2	−0.4	+0.2	−0.4	0.0	+0.3	0.0	+0.6		correction in milliohms

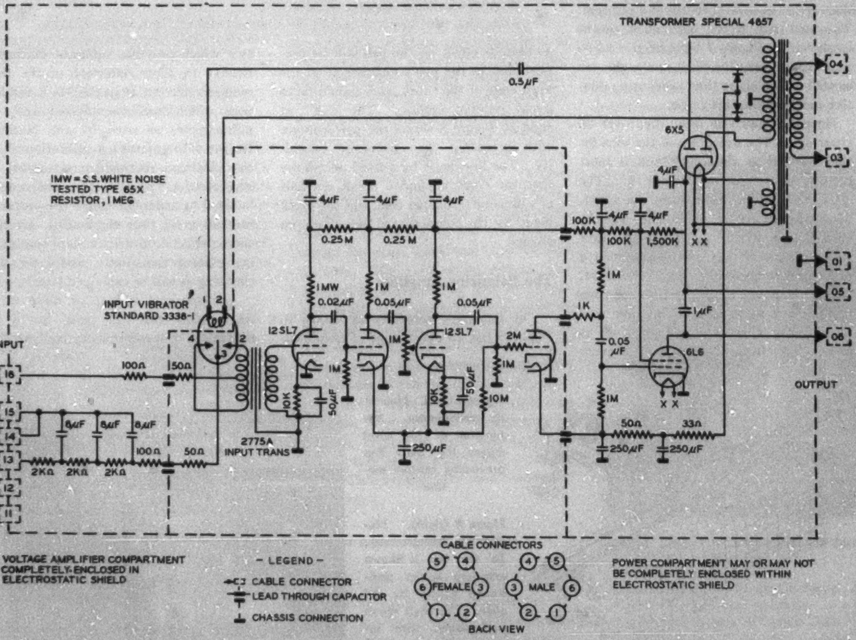

TRANSFORMER SPECIAL 4857
0.5μF
6X5
4μF
04
03
01
05
06
OUTPUT
IMW = S.S. WHITE NOISE TESTED TYPE 65X RESISTOR, I MEG
4μF
4μF
4μF
0.25 M
0.25 M
100K
4μF
4μF
100K
1,500K
X X
IMW
IM
IM
IM
0.02μF
0.05μF
0.05μF
IK
1μF
INPUT VIBRATOR STANDARD 3338-I
1
2
4
2
3
12SL7
12SL7
2M
0.05 μF
6L6
INPUT
16
15
14
13
12
11
100Ω
50Ω
IM
IM
IM
IM
10K
50μF
10K
50μF
10M
X X
8μF
8μF
8μF
100Ω
50Ω
2KΩ
2KΩ
2KΩ
2775A INPUT TRANS
250μF
50Ω
33Ω
250μF
250μF
VOLTAGE AMPLIFIER COMPARTMENT COMPLETELY ENCLOSED IN ELECTROSTATIC SHIELD
– LEGEND –
CABLE CONNECTOR
LEAD THROUGH CAPACITOR
CHASSIS CONNECTION
CABLE CONNECTORS
FEMALE
MALE
BACK VIEW
POWER COMPARTMENT MAY OR MAY NOT BE COMPLETELY ENCLOSED WITHIN ELECTROSTATIC SHIELD

Figure 10. Circuit of balancing amplifier. An extra stage provides ample amplification

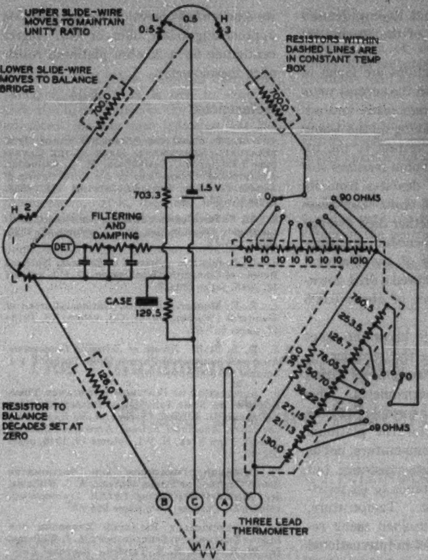

UPPER SLIDE-WIRE MOVES TO MAINTAIN UNITY RATIO
L
0.5
0.5
H
3
RESISTORS WITHIN DASHED LINES ARE IN CONSTANT TEMP BOX
LOWER SLIDE-WIRE MOVES TO BALANCE BRIDGE
700.0
700.0
703.3
1.5 V
0
90 OHMS
H 2
FILTERING AND DAMPING
I
DET
L
I
10 10 10 10 10 10 10 10 10
CASE
129.5
760.5
253.5
126.7
76.05
50.70
39.0
36.22
27.15
21.13
130.0
0
9 OHMS
126.0
RESISTOR TO BALANCE DECADES SET AT ZERO
B
C
A
THREE LEAD THERMOMETER

Figure 11 (above). Bridge circuit. The thermometer being in the same arm as the decades, a substitution type of measurement is made

Figure 12 (above). Interior of constant temperature box

Fifty-five degrees centigrade was selected for the control point because the temperature inside of the recorder may rise 15 degrees centigrade above a possible ambient of 40 degrees centigrade

detector and the network for filtering and damping just mentioned.

The effect of contact resistance in the 10-ohm decade is suppressed by a factor of five by placing it in a relatively high-valued ratio arm. The effect of contact resistance in the 1-ohm decade is suppressed by a factor of six or more by the use of a shunted type of decade as in Mueller's hand-balanced bridge.

The thermometer being in the same arm as the decades, a substitution type of measurement is made. This gives constant sensitivity, allows the use of a simple "Wagner" ground and avoids the need for checking and adjusting the ratio arms specifically. Any change in the ratio arms is indicated by a change in the zero reading and simply requires a new zero correction.

The critical resistors within the dashed lines are located in a constant temperature box. Figure 12 shows its interior.

Constancy to 0.001 ohm in 100 ohms or 1 part in 100,000 was achieved by temperature control to 1 degree centigrade and reduction of temperature dependence in the resistors to 1 part in 100,000 per degree centigrade. Fifty-five degrees centigrade was selected for the control point because the temperature inside of the recorder may rise 15 degrees centigrade above a possible ambient of 40 degrees centigrade. The critical resistors were made to have their minimum temperature dependence in the region of 55 degrees centigrade by making each resistor partly of manganin and partly of copper, in series combination.

The Slide-wire

Figure 13 shows the slide-wire and its drive. A single moulded disk carries both

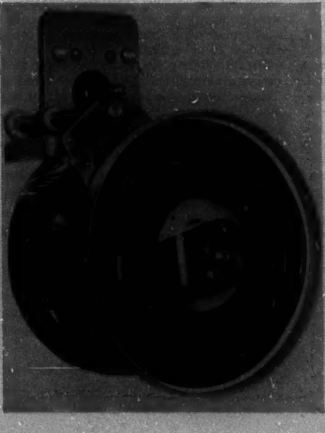


Figure 13 (left). Slide-wire and its drive

The slide-wire is continuous and so places no limit on the sensitivity of the recorder. It is housed in a magnetic shield of which only the back section is shown

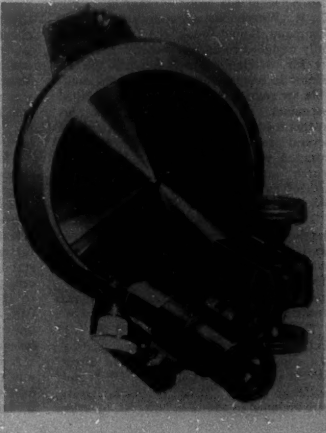


Figure 14 (right). Setup for precise checking

Lines on the disk fastened to the shaft are observed with reference to an index fastened to the slide-wire shield. A reticule in the microscope permits reading to 0.001 inch or 1/15,000th of the slide-wire travel of the slide-wires which appear in the bridge circuit. The disk is housed in a highly permeable magnetic shield of which only the back section is shown. Each slide-wire is continuous (not wound on a mandrel) and so places no limit on the sensitivity of the recorder.

Setup for Precise Checking

For ordinary use a sensitivity to about 1/1000th of the slide-wire span or 0.001 ohm is adequate, but for checking purposes more sensitivity was desired. The resistors in the recorder were intercompared by the so-called step-up method. In this method each 1-ohm step is compared to the slide-wire and each 10-ohm step is compared with the sum of the nine 1-ohm steps plus the slide-wire. Since the final calibration involves the sum of ten or more readings it was desirable to make each of these ten readings to one extra figure or to 0.0001 ohm.

It was found possible to do this with the setup shown in Figure 14. Lines on the disk, which is fastened to the slide-wire-contact shaft, are observed with reference to an index fastened to the slide-wire shield. A reticule in the microscope permits readings to 0.001 inch or 1/15,000th of the travel of the slide-wire contact. With the balancing amplifier set to optimum gain, readings were found to be reproducible to a value approaching 0.001 inch or 0.0001 ohm.

With these microscope readings it was possible to calculate errors and corrections for the 1-ohm and 10-ohm decades based on the average of the 10-ohm resistors.

These decade corrections are given in Table I. Apparently in the several years since these resistors were made and adjusted, the resistors making up the 1-ohm decade increased with respect to the 10-ohm resistors. The 10-ohm resistors still form a decade which deviates from its average by only 0.0006 ohm. The question might be asked whether their average had changed. Independent measurements with Standards of the National Bureau of Standards showed a drift downward in their resistance of about 0.005 per cent.

It is not necessary to know the exact amount of drift to use the recorder for precise temperature measurements with a platinum resistance thermometer. This is so since the Callendar formula[1] and the Van Dusen formula[2] for the platinum thermometer give the temperature, not as a function of its absolute resistance, but as the ratio of its resistance to its resistance at the ice point. Temperature, therefore, can be determined using resistance values measured in international ohms or absolute ohms or nonstandard ohms, provided only that the ice-point resistance is determined in terms of the same unit.

Summary

An attempt has been made to describe those features of a recorder which permit its unattended operation with a precision approaching 0.001 ohm or 0.01 degree centigrade for a 25-ohm platinum resistance thermometer.

References

1. On the Practical Measurement of Temperature: Experiments Made at the Cavendish Laboratory, Cambridge, H. L. Callendar. *Philosophical Transactions of the Royal Society of London* (London, England), volume 178, 1887, pages 161–230.

2. **M. S. Van Dusen.** *Journal*, American Chemical Society (Washington, D. C.), volume 47, 1925, pages 326–32.

3. **H. F. Stimson.** *Journal of Research*, National Bureau of Standards (Washington, D. C.), volume 42, 1948, pages 209–17.

4. **E. F. Mueller.** *Bulletin*, National Bureau of Standards (Washington, D. C.), volume 13, 1916–17, pages 547–61.

5. **D. R. Stull.** *Review of Scientific Instruments*, American Institute of Physics (New York, N. Y.), volume 16, 1945, pages 318–21.

6. Application of Platinum Resistance Thermometry to Some Industrial Physicochemical Problems, D. R. Stull. *Industrial and Engineering Chemistry, Analytical Edition*, American Chemical Society (New York, N. Y.), volume 18, 1946, pages 234–42.

7. Combined Thyratron and Tachometer Speed Control of Small Motors, A. J. Williams, Jr. *Electrical Engineering* (*AIEE Transactions*), volume 57, October 1938, pages 565–68.

8. Electronically Balanced Recorder for Flight Testing and Spectroscopy, A. J. Williams, Jr., W. R. Clark, R. E. Tarpley. *Electrical Engineering* (*AIEE Transactions*), volume 65, April 1946, pages 205–08.

9. Measuring Instrument, T. R. Harrison. United States Patent Number *1,927,402*.

10. Compensating Means for Dimensional Changes in a Recorder Chart, F. F. Dern, Jr., A. J. Williams, Jr. United States Patent Number *2,432,229*.

11. D-C Amplifier Stabilized for Zero and Gain, A. J. Williams, Jr., R. E. Tarpley, W. R. Clark. *AIEE Transactions*, volume 67, part 1, 1948, pages 47–57.

Discussion

C. R. Witschonke (American Cyanamid Company, Bound Brook, N.J.): The beautifully designed resistance recorder described by Mr. Williams greatly extends the accuracy in temperature measurements available for routine use by relatively unskilled personnel. Dr. Stull's original publications in this field interested us and subsequently several resistance recorders were constructed which have been applied to research and production materials. An article[2] has been printed briefly outlining one of these recorders.

Two different resistance recorders were developed. Both recorders used 10-degree-centigrade spans with 1/2-degree-centigrade overlap at each end and automatic range changing. The 124-ohm platinum resistance elements, for which these instruments were designed, were interchangeable.

The 24-range recorder, designed for laboratory research, used 5-stage amplification of a chopped d-c bridge signal and direct current on the heaters as in the Leeds and Northrup model. Tuned feedback circuits were included in addition, however, in order to give relatively narrow band-pass amplifier characteristics. This resulted in more than ample sensitivity for the 1.7-mil thermometer current operation. In fact, a change of 0.002 to 3 ohm (equivalent to 0.004 to 6 degrees centigrade) was detected easily by the recorder. The ingenious microscope technique described by Mr. Williams was not used so that it can be said that the stability was only as good as could be read on the engraved scale attached to the recorder, or about ±0.005 degree centigrade.

The second 4-decade recorder was designed for routine measurements. Only three stages of amplification of the a-c Wheatstone bridge signal were used. A thermometer current of 4 mils yielded sufficient signal for a temperature sensitivity of better than 0.01 degree centigrade. The stability was of the order of ±0.01 degree centigrade. The temperature increase due to the 4-mil current amounted to about 0.05 degree centigrade, but this effect was quite reproducible so that it was corrected for in the calibration of the instrument.

A novel feature of both of these recorders was that temperatures were read directly from the scale for all the decades. This was accomplished by changing the ratio arms for each decade in order to compensate for the slight nonlinearity in the temperature coefficient of platinum. Thus two resistances were accurately calibrated for each decade and replacement thermometers were constructed to have identical temperature coefficients and ice-point resistances within the desired accuracy of ±0.01 degree centigrade. Temperature coefficients were found to be amazingly reproducible when the highest purity platinum available was used and the duplication of ice-point resistances to 0.02 ohm has also proved to be practical. It is felt that the convenience of reading temperatures directly more than compensates for the complications involved. This is especially true when only a limited number of decades are involved. The method probably would be impractical for a span of more than 50 decades, however.

In the development of the recorders we were greatly aided by the Thwing-Albert Instrument Company and our electronics consultant, Mr. C. W. Warren.

References

1. See reference 5 of the paper.

2. Freezing Points in Determination of Product Purity, C. R. Witschonke. *Analytical*

Chemistry, American Chemical Society (New York, N. Y.), volume 24, 1952, page 350.

Daniel R. Stull (Dow Chemical Company, Midland, Mich.): It is well to examine periodically one's tools in the light of recent advances. As has been indicated, the International Temperature Scale from the boiling point of oxygen to the melting point of antimony is defined by international agreement in terms of a platinum resistance thermometer. It is, therefore, appropriate that precision measurement of such physical properties as freezing point, boiling point, vapor pressure, and so forth, be made with a platinum resistance thermometer. Previously, use of a platinum resistance thermometer tied a person to a measuring bridge. Use of skilled help to operate bridge dials probably does not utilize the individual's skill at anywhere near its maximum usefulness. The fully automatic recorder described here frees that individual to apply his talent and skill to problems of greater importance than that of seeking bridge balance and recording it.

A. J. Williams, Jr.: Dr. Stull's discussion explains very concisely the need for a recorder of the general type described in the paper.

Dr. Witschonke's discussion also indicates his appreciation of this need. His recorders were similar in many ways to the recorder described in the paper; for example, they had slide-wire spans of approximately 10 degrees centigrade and the recorder of greatest precision had direct current on the heaters of the critical amplifier tubes.

Dr. Witschonke has stated the important pros and cons on the direct-reading feature. Because we wished to cover the full range of the platinum resistance thermometer without risking the introduction of an additional error, we decided to build the simpler type of recorder which reads in resistance. The computation of temperature from resistance need not introduce any appreciable error.

The Fundamental Accuracy of Single-Phase 3-Wire Metering

E. C. WENTZ
MEMBER AIEE

A. J. PETZINGER
MEMBER AIEE

Synopsis: As a result of constantly increasing loads, 3-wire single-phase service is rapidly replacing the original 2-wire domestic service. This change has been accompanied by continuous improvement in the range and performance of the meters. In view of these developments, it is appropriate to analyze the bases upon which the fundamental correctness of 3-wire single-phase metering rests and which have been responsible for its universal acceptance for many years. The analysis in this paper shows that this acceptance is well founded, that the probable error of this form of metering can be calculated readily, and that it is usually negligible.

THE domestic consumption of electricity has increased constantly since the introduction of the a-c system. The past two decades have witnessed the general change from 2-wire to 3-wire service. While the 3-wire meter accounted for only 13 per cent of the single-phase watt-hour meters manufactured in 1925, this percentage had changed to 80 per cent in 1950. During this same 25-year period the average annual energy consumption per residential customer increased from 400 to 1,800 kilowatt-hours.

This period also has witnessed continuous improvement in performance and accuracy of the watt-hour meter. The overload capacity has been approximately tripled. Accuracy has been improved for load, voltage, and temperature variations. Meters have been protected against damage by lightning surges. Similarly, 3-wire current transformers also have been improved in accuracy and load range.

In view of these developments, an up-to-date analysis of the reasons for the fundamental accuracy of the single-element 3-wire metering circuit is appropriate. Application of the theory of probability in a rigorous manner shows why the accuracy of this method is well within commercial limits.

Outline of Method of Analysis

The approach of this paper is:

1. To calculate the error in terms of possible unbalances and circuit constants.
2. To determine the probability of all possible unbalances.
3. To evaluate the most probable error by a summation of the products of error and probability.
4. To evaluate the effect of simplifying assumptions made in analyzing the accuracy of registration.

Since the object is to analyze the fundamental accuracy of the circuit, effects, depending on the meter's calibration, load curve, balance of windings, voltage curve, and so forth, are not considered. (It is well known that the modern meter's performance is very good on these points.)

The Effect of Unbalanced Voltage on the Accuracy of a 3-Wire Meter

In accordance with Blondel's Theorem, two meter elements are required for accurate measurement of power in any 3-wire circuit. However, if the voltages from lines to neutral in a single-phase 3-wire circuit are equal, a single element with two current windings will meter the power correctly. If the voltages are slightly unbalanced, as is usually the case, very small errors result, providing the load also is unbalanced. Ordinarily, of course, unbalanced load is the only cause of unbalanced voltage.

Such a scheme is shown in Figure 1. If the loads from each line to neutral are equal, the voltages will be equal and the energy will be measured correctly. However, if the loads are not equal, the voltages will become unbalanced because of the unequal voltage drops in the transformer secondary and in the lines. The combination of unequal loads and unbalanced voltages causes slight errors.

The magnitude of the error is developed in Appendix I in terms of currents, voltages, and impedances. The error may be expressed as the product of half the unbalance in voltage times the power component of the neutral current. (Throughout the paper the dot product of two vectors such as $\mathbf{E}\cdot\mathbf{I}$ indicates $EI \cos \theta$ where θ is the angle between the vectors.)

$$\text{Error} = \left(\frac{\mathbf{E}_{10} - \mathbf{F}_{02}}{2}\right)\cdot\mathbf{I}_0$$

This is a general expression, independent of the number or size of the loads on the circuit, or the transformer and line impedances.

Paper 52-163, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted March 6, 1952; made available for printing April 17, 1952.

E. C. WENTZ is with the Westinghouse Electric Corporation, Sharon, Pa., and A. J. PETZINGER is with the Westinghouse Electric Corporation, Newark, N. J.

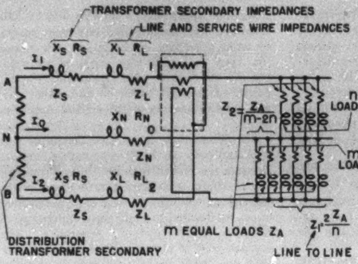


Figure 1. Circuit used for analyzing the effect on registration when a customer's load is unbalanced

This expression for error is of limited usefulness because the unbalance of load is not usually known. To facilitate the analysis, it is assumed for the moment that there is a single customer connected to the distribution transformer as shown in Figure 1. This customer has *m* equal loads, of which *n* are connected from one line to neutral. Appendix II shows that the percentage error may then be expressed as

$$\text{Per-cent error} = 100\,\frac{(R_S+R_L+2R_N)}{2R_A}\times\frac{(m-2n)^2}{m} \qquad (11)$$

$$= 100\,\frac{(R_S+R_L+2R_N)}{2R_2}\,\frac{W_1}{W_2} \qquad (16)$$

$$= \frac{100\,W_L}{W_1} \qquad (19)$$

where

R_S = one-half the transformer secondary resistance
R_L = resistance of one line wire and one service wire
R_N = neutral resistance
R_A = resistance of each load unit
m = total number of load units
n = number of load units connected from one line to neutral
$R_2 = R_A/(m-2n)$
W_1 = total power consumed
W_2 = power consumed by unbalanced portion of the load
W_L = total additional line and transformer losses caused by the unbalance

Equations 11, 16, and 19 all show that the error is plus. The reason for this is apparent from physical considerations. A load from one line to neutral reduces the voltage between that line and neutral because of impedance drops in the line transformer secondary. However, the voltage applied to the meter is the line-to-line voltage which has decreased only half as much on a percentage basis. The error is, therefore, plus for the condition of a single customer connected to the distribution transformer.

Equation 11 expresses the error in terms of the circuit resistances and the unbalance in the load. Since *m* represents the total number of load units, and *n* the number connected from one line to neutral, the number of balanced load units is represented by $2n$, and the number of unbalanced load units by $m-2n$.

It will be observed that for a single load unit of resistance R_A, the per-cent error is $100\,(R_S+R_L+2R_N)/2R_A$ since $m=1$ and $n=0$. This might be called the basic error and, as would be expected, depends on the ratio of circuit resistance to load resistance. For other values of *m* and *n* the basic error is multiplied by $(m-2n)^2/m$ or $[(m-2n)/m]\,(m-2n)$. $(m-2n)/m$ represents the portion of the load which is unbalanced and $m-2n$ the number of unbalanced load units.

Equation 16 expresses the error in terms of the ratio of circuit resistance to unbalanced load resistance times the ratio of unbalanced power to total power.

It is also apparent from equations 11 and 16, as would be expected, that the error is proportional to the ratio of transformer and line resistance to load resistance.

Equation 19 is very informative, since it shows that the error is always equal to the total additional losses in the line and transformer caused by the unbalanced portion of the load. For example, the transformer's secondary copper loss and line losses for a balanced load of 240 watts may total 0.2 watt. If an unbalanced load of 240 watts causes these losses to rise to 0.7 watt, the increased loss attributable to the unbalance is 0.5 watt. According to equation 19 the error in registration for the unbalanced condition therefore would be 0.5 watt or slightly less than 0.2 per cent. In other words, the meter measures the load power plus the increase in transformer and line losses caused by unbalance.

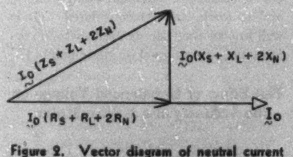


Figure 2. Vector diagram of neutral current and associated voltage drops

Probability of Unbalanced Load

Equations 11, 16, and 19 are sufficient to calculate the registration for any given load unbalance.

The next step is to calculate the probability of unbalance, assuming that, when a load is applied, there is an equal probability of its being connected to either line.

Appendix III shows that if there are *m* equal load units the probability of *n* units being connected from one line to neutral is $m!/2^m n!(m-n)!$ As an example, if there are 20 loads ($m=20$) the chances of complete unbalance ($n=0$ and $n=m$) and $1/2^{19}=1/524{,}288$ or not quite two in a million.

Probable Error

The probable error for a given unbalance is the error for that unbalance multiplied by the probability of that unbalance. The total most probable error is the summation of these probable errors for all possible unbalances. Equation 39 developed in Appendix IV shows that the total most probable error in per cent is $100\,(R_S+R_L+2R_N)/2R_A$

where

R_S = one-half the transformer secondary resistance
R_L = resistance of one line wire and one service wire
R_N = neutral resistance
R_A = resistance of each load unit

In other words, the probable percentage error is that which is caused by a single load unit being connected from line to neutral. Whether there are 10, 20, or 100 such load units to be connected, the probable percentage error remains that of a single unbalanced unit. This is somewhat surprising, since the maximum possible percentage error increases with the number of load units. However, the probability of encountering complete unbalance of a large number of units is so remote that the probable percentage error remains that caused by the unbalance of a single unit.

Another conclusion, again fairly obvious, is that the probable error is dependent on the ratio of transformer and

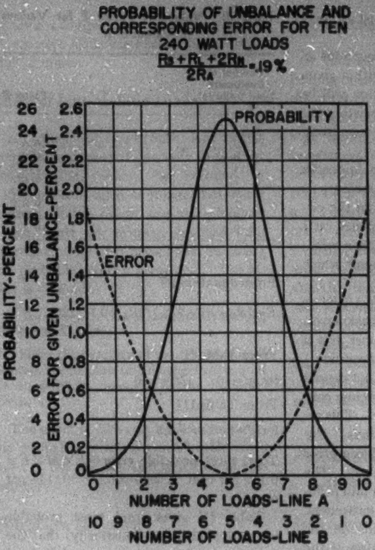


Figure 3 (left). Probability of unbalance and corresponding error

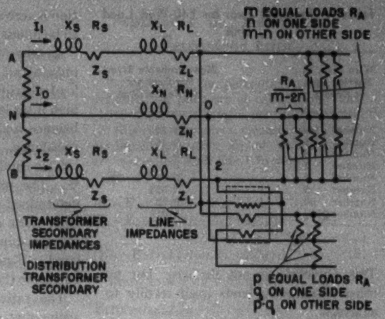


Figure 5 (right). Circuit used for analyzing the effect of other customers' unbalanced load on a given customer's registration

line resistance to load resistance. Low supply resistance (which is equivalent to high capacity) reduces errors. Similarly, the smaller the size of the assumed load unit, the smaller the errors.

Numerical Example

At this point it is well to consider some numerical values. The assumed size of the 2-wire load unit is not subject to precise calculation or prediction, since loads vary from electric clocks to ironers, a span of 2 to 1,500 watts. Units in the upper range, however, are relatively rare compared to the number of lamps and appliances in the 50- to 240-watt span. Assuming a basic load value of 240 watts or $R_A=60$ is probably pessimistic.

If the distribution transformer is rated at 10 kva, the copper loss in the secondary at rated load is probably approximately 90 watts.

$10{,}000/240=41.67$ amperes
$41.67^2\times 2R_S=90$
$R_S=90/3{,}470=0.026$ ohm

The resistance of number 4 wire is about 0.25 ohm per 1,000 feet. Assuming 150 feet of number 4 line wire, $R_L=R_N=0.038$ ohm. The service wire may be 75 feet of number 6 or $0.075\times 0.394=0.030$ ohm to be added to R_L and R_N. Substituting in equation 39: $R_S=0.026$, $R_L=0.068$, $R_N=0.068$, and $R_A=60$.

$$\frac{100(R_S+R_L+2R_N)}{2R_A}\text{ per cent}$$
$$=\frac{100(0.026+0.068+0.136)}{2\times 60}\text{ per cent}$$
$$=0.19\text{ per cent}$$

The probable error thus is 0.19 per cent.

Figures 3 and 4 illustrate the factors involved for the same impedance values but for a total of ten loads on the line ($m=10$). (While each variable is shown by a smooth curve for easier visualization, only the points corresponding to integral load values are significant.)

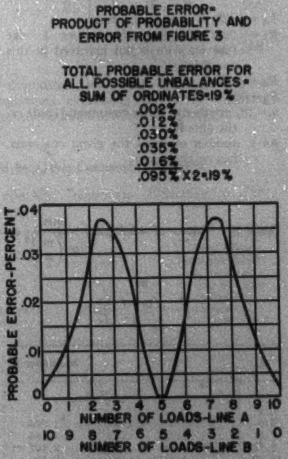


Figure 4. Probable error caused by unbalanced loads

The solid curve of Figure 3 shows the probability of a given number of loads being connected to one line. This curve is very similar to the "normal" or Gaussian distribution curve which frequently is used in quality control analysis. It is interesting to note that the chances are almost one out of four that the ten loads will be balanced. Also, the chances are almost two out of three (two-to-one odds) that ten loads will be split either five-five or six-four.

The dotted curve of Figure 3 shows the error in per cent for each condition of load distribution.

Figure 4 represents the product of the two curves of Figure 3. Therefore, the ordinate at each point represents the probable error due to that particular division of load. The sum of these ordinates is, therefore, the total most probable error for all possible unbalances. This sum equals 0.19 per cent as shown on Figure 4, thus checking equation 39.

If the basic load unit is assumed to be smaller than 240 watts, the probable error decreases proportionally. Table I shows the probable error caused by line and service wire resistances, assuming a basic load unit of 240 watts.

The transformer introduces a slight additional error, but in the numerical example the probable error caused by a 10-kva transformer is only 0.02 per cent so it can be disregarded for all practical purposes.

Effect of Unequal Loads

The solution presented involved the assumption that all load units were equal.

Table I. Probable Error for 240-Watt Load Unit

Wire Size, in Service Run or Feeder	Most Probable Error, Per Cent per 100 Feet of Run
Number 10	0.25
Number 8	0.16
Number 6	0.1
Number 4	0.06
Number 2	0.04
Number 0	0.025
Number 00	0.02
Number 0000	0.012

This obviously is not the case, but no other assumption lends itself readily to analysis.

However, it seems logical to assume that the probable error at any given time will not exceed the value calculated for a load unit equal to the largest unit in service at that time. For example, if the largest single load connected line to neutral is a 1,000-watt toaster, a probable error calculation based on all equal loads of this size is, if anything, pessimistic. Offsetting such a condition is the fact that for extended periods the largest single load in service is only 100 watts or less. Averaging the periods of service of the various loads, therefore, would probably give a load unit value of considerably less than the 240 watts assumed in the example.

Further indication that 240 watts is, if anything, a pessimistic value for each load unit is derived from annual residential kilowatt-hour consumption data. For the average residential consumption of 1,800 kwh per year, the total of all the loads of a residence average only 300 watts on a 16-hour-a-day basis.

Effect of Other Customers' Loads

The foregoing analysis is valid for the error of a given customer caused by his own unbalance. However, if there are a number of customers on the line, there is an additional error dependent on the load balance of the other customers.

Assuming, for the moment, that there are a number of customers on a line and that a given customer has a single load unit connected between line 1 and neutral, there are three possibilities regarding the rest of the load on the lines:

1. Remainder of load also heavier on line 1.
2. Remainder of load balanced.
3. Remainder of load heavier on line 2.

The effect of these loads on this customer's registration is to give plus errors in case 1, no error in case 2, and minus errors in case 3. The most probable error caused by the other loads on his registration is zero since the effect is just as likely to be minus as plus.

In other words, if any customer adds an unbalanced load to a circuit with a given load distribution, his registration will be high if his load aggravates an existing unbalance, and low if it decreases an existing unbalance.

This condition is analyzed in Appendix V. The probable voltage unbalance is determined for any total load on the circuit. The effect of this unbalance on a given customer's registration is then calculated. Figure 5 shows the circuit which applies to this analysis.

The ratio $E=(R_S+R_L+2R_N)/2R_A$ is again what might be called a basic error. Assuming that a given customer has a single load unit connected between line 1 and neutral, the basic error is that caused by another single load unit being connected by some other customer. This error may be either plus or minus. For more general conditions this basic error must be multiplied by two factors. The first, M, depends on the number and probable distribution of other customers' loads. The second, F, depends on the number and probable distribution of the given customer's loads. The total most probable error caused by other customers' loads is $\pm E\times M\times F$ as given by equation 50 in Appendix V.

This equation is now applied to an example. The same values are assumed as in the previous example (except that the customer's service wire resistance is omitted since other customer's loads are not carried by it).

$R_S=0.026$, 10-kva transformer
$R_L=0.038$, 150 feet of number 4 wire (service wire is not involved in this condition)
$R_N=0.038$
$R_A=60$
$m=10$, number of other customers' loads on the circuit
$p=3$, number of loads the given customer has on the circuit

Table II. Values of Multiplier $\dfrac{m\,m!}{2^m\left(\dfrac{m}{2}!\right)^2}$ or $\dfrac{m!}{2^{m-1}\left(\dfrac{m-1}{2}!\right)^2}$

m	Multiplier M
1,2	1
3,4	1.5
5,6	1.875
7,8	2.188
9,10	2.461
19,20	3.524
39,40	5.015
99,100	7.959

Table III. Values of Factor F for Various Values of p

Number of Customers Load Units (p)	Factor F
1	1
2	0.5
3	0.5
4	0.375
5	0.375
6	0.312
7	0.312
8	0.273
9	0.273

From equation 39

$$E=(R_S+R_L+2R_N)/2R_A=0.14/120$$
$$=0.117 \text{ per cent}$$

From Table II

for $m=10$ $M=2.46$

From Table III

for $p=3$ $F=0.5$

Total most probable error $=E\times M\times F=$ 0.117 per cent $\times$ 2.46 $\times 0.5=\pm 0.144$ per cent.

Table IV gives total most probable errors assuming (pessimistically) that the transformer is operating at full load $(p+m)=42$.

Table IV shows that, if the customer has more than one load unit connected, the probable error is reduced rapidly because the error caused by other customers applies only to the unbalanced portion of his load. For example, the probable error is ±0.3 per cent if the customer has two loads connected, compared to ±0.6 per cent if he has only one load connected.

If the basic load unit is 100 watts instead of 240 watts, the probable error is reduced to about two-thirds of the values just given for given values of total load. It should be noted therefore that this component of the error is rarely as great as 0.3 per cent and tends to average zero.

Effect of Transformer and Line Reactance

One other factor which has been neglected in the foregoing analysis is reactance of transformer and line. Resistive loads cause voltage drops in line and transformer resistances. Likewise, the reactive component of the load causes a voltage drop of the same phase position in the transformer and line reactances. The effect is similar to that described in the foregoing and may be analyzed in the same way. Transformer reactances are normally approximately the same as the resistance, while line reactance is approximately 0.15 ohm per 1,000 feet. The net effect for most conditions is, therefore, somewhat comparable. The errors caused by reactive loads are equally likely to be plus or minus.

Effect of Line-to-Line Loads

All calculations and conclusions to this point have neglected one very important consideration, namely, the existence of 240-volt loads, such as ranges, hot water heaters, clothes driers, and some motors. These loads are always balanced and, therefore, measured accurately, reducing the probable errors a great deal. The effect is twofold:

1. Total registration is increased, reducing the percentage errors.
2. Since the capacity of the transformer is limited, the possible unbalanced load cannot be as great as assumed in the examples given.

The effect of 240-volt loads in minimizing probable errors is considerable since most of the heavier load units are in this category.

Polyphase Metering

The foregoing analyses of the accuracy of the 3-wire single-phase meter are applicable to a large extent to the 3-wire elements of polyphase meters. Two meters in common use incorporate such elements, namely, the 2-element meter for 3-phase 4-wire delta circuits, and the "$2^1/_2$ element" meter for 3-phase 4-wire circuits. The general assumption of balanced voltage also applies to these meters. The same method of analysis can be used and the conclusions may safely be expected to be comparable; that is, errors probably will not exceed tenths of a per cent. For polyphase service a larger percentage of the load normally will be balanced than for single phase.

Conclusions

A thorough analysis of the probable loading of single-phase 3-wire circuits and the effects on voltage balance shows that the registration of the single-element 3-wire meter is extremely accurate. For unbalances created by a customer, the probable error is of the order of +0.2 per cent even under pessimistic assumptions as to load size. Probable unbalance of other customers' loads, again with extremely pessimistic assumptions, results in an error of the same order of magnitude and this tends to average zero. In general, it seems safe to conclude that over a period of a month, the errors caused by unbalanced loads will be 0.2 per cent or less. While such high accuracy has generally been assumed, the analyses provide a sound basis for this conclusion.

Appendix I. Calculation of Error Caused by Unbalanced Load

$$\text{True power}=\mathbf{E}_{10}\cdot\mathbf{I}_1+\mathbf{E}_{20}\cdot\mathbf{I}_2 \qquad (1)$$

$$\text{Measured power}=\mathbf{E}_{12}\cdot(\mathbf{I}_1-\mathbf{I}_2)/2 \qquad (2)$$

Error = true power − measured power

$$=\mathbf{E}_{10}\cdot\mathbf{I}_1+\mathbf{E}_{20}\cdot\mathbf{I}_2-\mathbf{E}_{12}\cdot(\mathbf{I}_1-\mathbf{I}_2)/2 \qquad (3)$$

But $\mathbf{I}_1+\mathbf{I}_2+\mathbf{I}_0=0$, so $\mathbf{I}_2=-\mathbf{I}_1-\mathbf{I}_0$. Also $\mathbf{E}_{12}=\mathbf{E}_{10}-\mathbf{E}_{20}$. Substituting these values in equation 3

$$\text{Error}=\mathbf{E}_{10}\cdot\mathbf{I}_1-\mathbf{E}_{20}\cdot\mathbf{I}_1-\mathbf{E}_{20}\cdot\mathbf{I}_0-\mathbf{E}_{10}\cdot\left(\frac{2\mathbf{I}_1+\mathbf{I}_0}{2}\right)+\mathbf{E}_{20}\cdot\left(\frac{2\mathbf{I}_1+\mathbf{I}_0}{2}\right)=-\mathbf{E}_{10}\cdot\left(\frac{\mathbf{I}_0}{2}\right)-\mathbf{E}_{20}\cdot\left(\frac{\mathbf{I}_0}{2}\right)=-(\mathbf{E}_{10}-\mathbf{E}_{02})\cdot\frac{\mathbf{I}_0}{2} \qquad (4A)$$

$$=-\left(\frac{\mathbf{E}_{10}-\mathbf{E}_{02}}{2}\right)\cdot\mathbf{I}_0 \qquad (4B)$$

Equations 4(A) and 4(B) are general equations expressing the error resulting from unbalance of load and voltage.

Table IV. Most Probable Errors Caused by Other Customers' Unbalance

$E=(R_S+R_L+2R_N)/2R_A=0.117$ Per Cent. $(p+m)=42$, Transformer Operating at Full Load. $R_A=60$

Number of Given Customers Load Units (p)	Number of Other Customers Load Units (m)	Multiplier M	Factor F	Total Most Probable Error in Per Cent Caused by Other Customers' Unbalance = E×M×F
1	41	5.14	1	±0.60
2	40	5.02	0.5	±0.30
3	39	5.02	0.5	±0.30
4	38	4.89	0.375	±0.21
5	37	4.89	0.375	±0.21
6	36	4.75	0.312	±0.17
7	35	4.75	0.312	±0.17
8	34	4.61	0.273	±0.15
9	33	4.61	0.273	±0.15
10	32	4.47	0.246	±0.13
11	31	4.47	0.246	±0.13
12	30	4.34	0.221	±0.11

Appendix II. Determining Percentage Error

To determine the percentage error, it is assumed, for purposes of calculation, that: (1) a single customer is on the line; and (2) this customer's load consists of m identical impedance units, Z_A, of which n units are connected from line 1 to neutral.

From Figure 1

$$\mathbf{E}_{10}=\mathbf{E}_{AN}-\mathbf{I}_1(Z_S+Z_L)+\mathbf{I}_0Z_N$$

$$\mathbf{E}_{20}=\mathbf{E}_{BN}-\mathbf{I}_2(Z_S+Z_L)+\mathbf{I}_0Z_N=-E_{AN}+\mathbf{I}_1(Z_S+Z_L)+I_0(Z_S+Z_L+Z_N)$$

$$(\mathbf{E}_{10}-\mathbf{E}_{02})=\mathbf{I}_0(Z_S+Z_L+2Z_N) \qquad (5)$$

From equations 4(B) and 5

$$\text{Error}=\mathbf{I}_0\cdot\mathbf{I}_0(Z_S+Z_L+2Z_N)/2=I_0^2(R_S+R_L+2R_N)/2 \text{ from Figure 2} \qquad (6)$$

$$\text{Per-cent error}=\frac{100I_0^2(R_S+R_L+2R_N)/2}{\mathbf{E}_{12}\cdot(\mathbf{I}_1-\mathbf{I}_2)/2}=\frac{100I_0^2(R_S+R_L+2R_N)/2}{\mathbf{E}_{12}\cdot(2I_1+I_0)/2} \qquad (7)$$

$$\text{But } \mathbf{I}_1=\frac{E_{12}}{Z_1} \text{ and } \mathbf{E}_{12}\cdot 2\mathbf{I}_1=\frac{2E_{12}^2R_1}{Z_1^2} \qquad (8)$$

$$\mathbf{I}_0=\frac{\mathbf{E}_{12}}{2Z_2} \text{ and } \mathbf{E}_{12}\cdot\mathbf{I}_0\approx\frac{E_{12}^2R_2}{2Z_2^2} \qquad (9)$$

Substituting these values in equation 7

$$\text{Per-cent error}=\frac{100(R_S+R_L+2R_N)/2}{4R_1(Z_2/Z_1)^2+R_2}$$

If all load units are equal, $R_2/Z_2=R_1/Z_1$ or $R_2=R_1(Z_2/Z_1)$.

$$\text{Per-cent error}=\frac{100(R_S+R_L+2R_N)/2}{4R_1\left[\left(\frac{Z_2}{Z_1}\right)^2+\frac{Z_2}{4Z_1}\right]} \qquad (10)$$

From Figure 1

$$\frac{Z_2}{Z_1}=\frac{Z_A}{m-2n}\,\frac{n}{2Z_A}=\frac{n}{2(m-2n)}$$

$$R_1=2R_A/n$$

Substituting these values in equation 10 and simplifying

Per-cent error =

$$\frac{100(R_S+R_L+2R_N)/2}{2R_A\left[\left(\frac{n}{m-2n}\right)^2+\frac{n}{2(m-2n)}\right]}=\frac{100(R_S+R_L+2R_N)(m-2n)^2}{2mR_A} \qquad (11)$$

This may be expressed in several alternative forms.

$$\text{Let } W_1=m\left(\frac{E_{12}}{2Z_A}\right)^2R_A\approx\text{total power} \qquad (12)$$

$$W_2\approx(m-2n)\left(\frac{E_{12}}{2Z_A}\right)^2R_A \text{ total unbalanced power} \qquad (13)$$

Then (W_1-W_2) = balanced power

From Figure 1

$$R_2\approx R_A/(m-2n) \text{ and } Z_2=Z_A/(m-2n) \qquad (14)$$

From equations 12 and 13

$$W_2/W_1=(m-2n)/m \qquad (15)$$

Substituting equations 14 and 15 in equation 11

$$\text{Per-cent error}=\frac{100(R_S+R_L+2R_N)}{2R_2}\,\frac{W_2}{W_1} \qquad (16)$$

From equations 9, 13, and 14

$$W_2=I_0^2R_2 \qquad (17)$$

Substituting equation 17 in 16

$$\text{Per-cent error}=\frac{100(R_S+R_L+2R_N)}{2}\,\frac{I_0^2}{W_1} \qquad (18)$$

The transformer and line losses may be compared for a given load under balanced and unbalanced conditions.

Let

W_B = losses caused when a given load is balanced

W_U = losses caused when the same load is unbalanced

$W_L=W_U-W_B$ = additional loss caused by the fact that the load is unbalanced

Then $W_B=\left(\frac{E_{12}}{4Z_A}\right)^2(2R_S+2R_L)$

$$W_U=\left(\frac{E_{10}}{Z_A}\right)^2(R_S+R_L+R_N)$$

where $\frac{E_{12}^2}{4Z_A}\approx\frac{E_{10}^2}{Z_A}$ (equal power under balanced and unbalanced conditions)

$$W_L=W_U-W_B=\left(\frac{E_{10}}{Z_A}\right)^2(R_S+R_L+R_N)-\left(\frac{E_{12}}{4Z_A}\right)^2(2R_S+2R_L)$$

But $\frac{E_{10}}{Z_A}=I_0$ and $E_{12}\approx 2E_{10}$

So $W_L=I_0^2(R_S+R_L+R_N)-\frac{I_0^2}{4}(2R_S+2R_L)$

$$W_L=I_0^2\left(\frac{R_S+R_L+2R_N}{2}\right)$$

Substituting in equation 18

$$\text{Per-cent error}=100\,\frac{W_L}{W_1} \qquad (19)$$

In other words, the increase in registration caused by the unbalanced load is equal to the additional transformer and line losses caused by the fact that a portion of the load is unbalanced.

Appendix III. Probability of Connecting *n* Units of Load Between One Line and Neutral when the Total Number of Load Units = *m*

The total number of combinations of m units taken n at a time $=C(m,n)=\frac{m!}{n!(m-n)!}$

As an example, $n=0$ and $n=m$ represent complete unbalance, one equivalent to connecting all the loads on one line and the other equivalent to connecting all the loads on the other line.

$$C(m,0)=\frac{m!}{0!(m!)}=1$$

$$C(m,m)=\frac{m!}{m!(0!)}=1$$

The total possible number of loading combinations is the summation of the values for all possible values of n or

$$\sum_{n=0}^{m}C(m,n)=\sum_{n=0}^{m}\frac{m!}{n!(m-n)!} \qquad (20)$$

The binomial theorem states

$$(a+b)^m=\sum_{n=0}^{m}C(m,n)a^{m-n}b^n$$

$$(1+1)^m=\sum_{n=0}^{n=m}C(m,n)=2^m$$

The probability of the existence of any given combination of loading is, therefore, the ratio of the number of combinations giving this loading to the total possible combinations

P_n = probability of n loads on one specified side $=\frac{C(m,n)}{2^m}=\frac{m!}{2^m n!(m-n)!}$ (21)

Appendix IV. Calculation of Total Most Probable Error

From Appendix II, equation 11, the error is given by

Per-cent error

$$=\frac{100(R_S+R_L+2R_N)(m-2n)^2}{2mR_A} \qquad (11)$$

and the probability of n loads being connected one one side from Appendix III, equation 21, is

$$P_n=\frac{m!}{2^m n!(m-n)!} \qquad (21)$$

so that the probable error for any loading n is the product of equations 11 and 21.

Probable error (per cent) due to any given number n of loads on one line $n=0$ to $n=m$

$$=\frac{100(R_S+R_L+2R_N)(m-2n)^2m!}{2mR_A2^mn!(m-n)!} \qquad (22)$$

The total most probable error will be the summation of all the probable errors from $n=0$ to $n=m$.

Total most probable error, per cent

$$=\frac{100\,(R_S+R_L+2R_N)}{2mR_A}\,\frac{m!}{2^m}\times\sum_{n=0}^{m}\frac{(m-2n)^2}{n!(m-n)!} \qquad (23)$$

Total most probable error, per cent

$$=\frac{100(R_S+R_L+2R_N)}{2mR_A}\,\frac{m!}{2^m}\times\left[\sum_{n=0}^{m}\frac{m^2}{n!(m-n)!}-\sum_{n=0}^{m}\frac{4nm}{n!(m-n)!}+\sum_{n=0}^{m}\frac{4n^2}{n!(m-n)!}\right] \qquad (24)$$

This can be simplified by using the Binomial Theorem to obtain values for the summations inside the bracket. In the following development, equations 25 and 26 simplify the first summation, equations 27 to 34 the second summation, and equations 35 to 38 the third summation.

$$(a+b)^m=\sum_{n=0}^{m}\frac{m!a^{(m-n)}b^n}{n!(m-n)!}$$

If $a=b=1$, then

$$2^m=\sum_{n=0}^{m}\frac{m!}{n!(m-n)!} \qquad (25)$$

Multiplying by $m^2/m!$

$$\sum_{n=0}^{m}\frac{m^2}{n!(m-n)!}=\frac{m^22^m}{m!} \qquad (26)$$

Replacing m by $(m-1)$ in equation 25 gives

$$2^{m-1}=\sum_{n=0}^{m-1}\frac{(m-1)!}{n!(m-1-n)!}=\sum_{n=0}^{m-1}\frac{(m-1)!(m-n)}{n!(m-n)!} \qquad (27)$$

$$=\sum_{n=0}^{m-1}\frac{m!}{n!(m-n)!}-\sum_{n=0}^{m-1}\frac{(m-1)!n}{n!(m-n)!} \qquad (28)$$

The upper limit may be extended to m by evaluating the term for $n=m$ and correcting for its value.

$$\text{If } n=m,\text{ then }\frac{m!}{n!(m-n)!}=1 \qquad (29)$$

so that $\sum_{n=0}^{m-1}\frac{m!}{n!(m-n)!}$

$$=\left[\sum_{n=0}^{m}\frac{m!}{n!(m-n)!}\right]-1 \qquad (30)$$

$$\text{And if } n=m,\text{ then }\frac{(m-1)!n}{n!(m-n)!}=\frac{m!}{m!0!}=1 \qquad (31)$$

so that $\sum_{n=0}^{m-1}\frac{(m-1)!n}{n!(m-n)!}$

$$=\left[\sum_{n=0}^{m}\frac{(m-1)!n}{n!(m-n)!}\right]-1 \quad (32)$$

Substituting equations 30 and 32 in 28

$$2^{m-1}=\left[\sum_{n=0}^{m}\frac{m!}{n!(m-n)!}\right]-1-\left[\sum_{n=0}^{m}\frac{(m-1)!n}{n!(m-n)!}\right]+1$$

Substituting equation 25 for the first term, this can be rewritten as

$$2^{m-1}=2^m-\sum_{n=0}^{m}\frac{(m-1)!n}{n!(m-n)!}$$

$$\sum_{n=0}^{m}\frac{(m-1)!n}{n!(m-n)!}=2^m-2^{m-1}=2^{m-1} \quad (33)$$

Multiplying 33 by $4m/(m-1)!$ gives

$$\sum_{n=0}^{m}\frac{4mn}{n!(m-n)!}=\frac{4m2^{m-1}}{(m-1)!}=\frac{2m^2 2^m}{m!} \quad (34)$$

Similarly, replacing m by $m-2$ in equation 25 gives

$$2^{m-2}=\sum_{n=0}^{m-2}\frac{(m-2)!}{n!(m-2-n)!}=\sum_{n=0}^{m-2}\frac{(m-2)!(m-n)(m-1-n)}{n!(m-n)!} \quad (35)$$

If $n=m-1, \quad m-1-n=0$
If $n=m, \quad m-n=0$

In either case equation 35 would be zero for $n=m-1$ or $n=m$ so that the upper limit of 35 can be extended to $n=m$ without changing its value. Expanding 35

$$2^{m-2}=\sum_{n=0}^{m}\frac{(m-2)![m(m-1)-n(m-1)-nm+n^2]}{n!(m-n)!} \quad (36)$$

$$2^{(m-2)}=\sum_{n=0}^{m}\left(\frac{m!}{n!(m-n)!}-\frac{(m-1)!n}{n!(m-n)!}-\frac{(m-2)!mn}{n!(m-n)!}+\frac{n^2(m-2)!}{n!(m-n)!}\right)$$

Substituting equations 25 and 33 for the first two terms give

$$2^{(m-2)}=2^m-2^{m-1}-\sum_{n=0}^{m}\left(\frac{(m-1)!mn}{n!(m-n)!(m-1)}-\frac{n^2(m-2)!}{n!(m-n)!}\right)$$

But $\sum_{n=0}^{m}\frac{(m-1)!mn}{n!(m-n)!(m-1)}=\frac{m}{(m-1)}\times$

$$\sum_{n=0}^{m}\frac{(m-1)!n}{n!(m-n)!}=\frac{m}{m-1}2^{(m-1)}$$

from equation 33

So $2^{m-2}=2^m-2^{m-1}-\frac{m}{m-1}2^{m-1}+\sum_{n=0}^{m}\frac{n^2(m-2)!}{n!(m-n)!}$

Therefore

$$\sum_{n=0}^{m}\frac{n^2(m-2)!}{n!(m-n)!}=2^{m-2}-2^m+2^{m-1}+\frac{m}{m-1}2^{m-1}=\frac{2^m}{4}-2^m+\frac{2^m}{2}+\frac{m2^m}{2(m-1)}=\frac{2^m}{4}\left(\frac{m+1}{m-1}\right) \quad (37)$$

Dividing equation 37 by $(m-2)!$ and multiplying by 4

$$\sum_{n=0}^{m}\frac{4n^2}{n!(m-n)!}=\frac{(m+1)}{(m-1)!}2^m=\frac{(m^2+m)2^m}{m!} \quad (38)$$

Equation 24 then can be rewritten using equations 26, 34, and 38 in equation 24 giving

Total most probable error, per cent

$$=\frac{100(R_S+R_L+2R_N)}{2mR_A}(m^3-2m^2+m^2+m)$$

$$=\frac{100[(R_S+R_L)+2R_N]}{2R_A} \quad (39)$$

Appendix V. Determination of Probable Error Caused by Other Customers' Loads (See Figure 5)

Excluding the effect of a single load R_A of the customer under consideration, it follows from Appendix II that the voltage error at the meter in per cent is

$$\frac{100\mathbf{I}_0(R_S+R_L+2R_N)/2}{(\mathbf{E}_{10}+\mathbf{E}_{02})/2} \quad (40)$$

$$\mathbf{I}_0=\mathbf{E}_{02}(m-2n)/R_A \text{ where } E_{02}\approx\frac{E_{10}+E_{02}}{2} \quad (41)$$

Then from equations 40 and 41 voltage unbalance error in per cent $=\frac{100R_S+R_L+2R_N}{2}\,\frac{m-2n}{R_A}$ approximately.

From equation 21, Appendix III, the probability of n loads on one specified side is $\frac{m!}{2^m n!(m-n)!}$ and the probability of n loads on either side is therefore

$$\frac{m!}{2^{m-1}n!(m-n)!} \quad (42)$$

The probable per-cent registration error is therefore

$$\frac{100(R_S+R_L+R_N)}{2R_A}\,\frac{m!}{2^{m-1}}\sum_{n=0}^{m/2}\frac{m-2n}{n!(m-n)!} \text{ for even values of } m \quad (43A)$$

$$\frac{100(R_S+R_L+2R_N)}{2R_A}\,\frac{m!}{2^{m-1}}\sum_{n=0}^{(m-1)/2}\frac{m-2n}{n!(m-n)!} \text{ for odd values of } m \quad (43B)$$

Both equations 42 and 43 would have probable values of zero if summed over the range $n=0$ to $n=m$, indicating that the registration error is equally likely to be plus or minus. When summed over positive values of $(m-2n)$ the probable magnitude of the error is indicated. The upper summation limits are different for odd and even values of m since n must be an integer.

Expansion of both summations gives

$$\frac{m!}{2^{m-1}}\sum_{n=0}^{m/2 \text{ or } (m-1)/2}\frac{m-2n}{n!(m-n)!}=1\times\frac{3}{2}\times\frac{5}{4}\times\frac{7}{6}\times\frac{9}{8} \text{ to } \left\{\begin{matrix}m/2 \text{ terms for even values of } m\\(m+1)/2 \text{ terms for odd values of } m\end{matrix}\right\} \quad (44)$$

By inserting the factors $\frac{2}{2}, \frac{4}{4}, \frac{6}{6}\ldots$ in alternate terms this expansion can be broken down.

$$=\left(1\times\frac{3}{2}\times\frac{5}{4}\times\frac{7}{6}\ldots\right)$$

$$=\left(1\times\frac{2}{2}\times\frac{3}{2}\times\frac{4}{4}\times\frac{5}{4}\times\frac{6}{6}\times\frac{7}{6}\ldots\right) \text{ to } m \text{ terms}$$

$$=\begin{cases}\dfrac{mm!}{(2^2\times4^2\times6^2\ldots m^2)} & \text{for even values of } m\\[2ex]\dfrac{m!}{[2^2\times4^2\times6^2\ldots(m-1)^2]} & \text{for odd values of } m\end{cases}$$

$$=\begin{cases}\dfrac{mm!}{2^m\left(\frac{m}{2}!\right)^2} & \text{for even values of } m \quad (45A)\\[2ex]\dfrac{m!}{2^{m-1}\left(\frac{m-1}{2}!\right)^2} & \text{for odd values of } m \quad (45B)\end{cases}$$

So the probable percentage error, either plus or minus, is

$$\frac{100(R_S+R_L+R_N)}{2R_A}\times\begin{cases}\dfrac{mm!}{2^m\left(\frac{m}{2}!\right)^2} & \text{for even values of } m \quad (46A)\\[2ex]\dfrac{m!}{2^{m-1}\left(\frac{m-1}{2}!\right)^2} & \text{for odd values of } m \quad (46B)\end{cases}$$

It will be observed that the multiplier for even values of m is the same as that for the next lower odd value of m.

A more convenient formula for determining the value of the multiplier can be developed with the use of Stirling's Formula

$$\sqrt{2k\pi}\left(\frac{k}{e}\right)^k < k! < \sqrt{2k\pi}\left(\frac{k}{e}\right)^k\left(1+\frac{1}{12k-1}\right) \tag{47}$$

Substituting the left-hand term of equation 47 in equation 45(A) and letting $m=k$

$$M=\frac{mm!}{2^m\left(\frac{m}{2}!\right)^2}=\frac{m\sqrt{2m\pi}\left(\frac{m}{e}\right)^m}{2^m\frac{2m\pi}{2}\left(\frac{m}{2e}\right)^m}$$

$$=\sqrt{2m/\pi}\text{ for even values of }m \tag{48A}$$

This is about 4 per cent high at $m=6$, less than 1 per cent high at $m=40$.

The use of the correction factor $[1+1/(12k-1)]$ on the right-hand side of equation 47 enables almost exact results to be obtained.

Applying this factor to numerator and denominator of 45(A) gives

$$M=\sqrt{2m/\pi}\left[\frac{1+\frac{1}{12m-1}}{\left(1+\frac{1}{6m-1}\right)^2}\right]$$

$$=\sqrt{2m/\pi}\left[\frac{\frac{12m}{12m-1}}{\left(\frac{6m}{6m-1}\right)^2}\right]$$

$$=\sqrt{2m/\pi}\left(\frac{12m}{12m-1}\right)\left(\frac{36m^2-12m+1}{36m^2}\right)$$

$$M=\sqrt{2m/\pi}\,\frac{36m^2-12m+1}{(36m^2-3m)}\approx$$

$$\sqrt{2m/\pi}\left(1-\frac{1}{4m}\right) \tag{48B}$$

Using this equation, excellent results are obtained for all even values of m. (Odd values of m give the same multiplier as the next higher even value.)

Table II applies when the customer under consideration has a single unbalanced load unit connected to the system. If this customer has a number of load units in service, the multiplier becomes smaller. This is so because the error caused by other customers' unbalances is a smaller part of his total registration. As given in equation 42, the probability of connecting q loads on either one line or the other is $p!/[2^{p-1}q!(p-q)!]$ where p is the total number of loads. This value must be multiplied by $(p-2q)/p$, the factor representing the unbalanced portion of the load. The additional factor to be applied to Table II is, therefore, the summation

$$\sum_{q=0}^{p/2}\frac{p-2q}{p}\,\frac{p!}{2^{p-1}q!(p-q)!} \tag{49}$$

for even values of p.

This can be rewritten as $\frac{1}{p}\,\frac{p!}{2^{p-1}}\sum_{q=0}^{p/2}\times\frac{p-2q}{q!(p-q)!}$ which is seen to be M/p where M is the multiplier in Table II. (Odd values of p are obviously handled in the same way as in the summations involving m and n.) Table III gives the values of factor F for various values of p.

The total most probable error caused by other customers' unbalance is, therefore, equal to the product of E, M, and F.

$$\text{Total most probable error}=E\times M\times F \tag{50}$$

where $E=\frac{100(R_S+R_L+2R_N)}{2R_A}$ the same as equation 39

M = multiplier given in Table II for any number (m) of other customers' basic load units connected to the same transformer

F = multiplier given in Table III for any number (p) of basic load units which the customer himself has connected.

Discussion

F. H. Busch (General Electric Company, Lynn, Mass.): E. C. Wentz and A. J. Petzinger's paper is a very thorough and commendable analysis of the fundamental accuracy of single-phase 3-wire metering. Combining the probability of unbalance with the error due to unbalance provides a very practical and clear method of stating the total possible error for all possible unbalances.

In commenting on the application of their analysis of the 3-wire single-phase meter to the polyphase meter with 3-wire elements, the authors state that for the 2-element meter for 3-phase 4-wire delta circuits, and for the "$2^1/_2$ element" meter for 3-phase 4-wire Y circuits, "The same method of analysis can be used, and the conclusions may safely be expected to be comparable; that is, errors probably will not exceed tenths of a per cent."

It should be pointed out that this conclusion overlooks one quite important consideration in the particular case of the $2^1/_2$ element meter for 3-phase 4-wire Y circuits. For this particular meter, the largest possibility for error arises when the line voltages are unbalanced. Such unbalance can be due to unbalanced loads on the circuit being metered, and the amount and effect probably will be small, just as analyzed by the authors. However, the unbalance also can be due to unbalanced voltage on the 3-phase distribution system supplying the transformer bank being metered. Significant errors may result even though the metered load is well balanced. For example, a 1-per-cent voltage unbalance may cause approximately 0.3% error for a balanced load on the meter.

Extra caution must be exercised to insure satisfactory voltage balance of the distribution system in the application of the $2^1/_2$ element meter. This unbalance of primary voltage supply may arise in a particular distribution system due to unequal single-phase branch loads supplying residential areas and may continue to exist for long periods of time. Such an unbalanced load, particularly for rural distribution systems, may be a definite part of the load growth plan, and, therefore, not lend itself to the probability treatment used by the authors.

E. C. Wentz and **A. J. Petzinger:** The only reason that the voltages in a 3-phase 4-wire system should be unbalanced is that someone somewhere on the system is overloading one phase. This could, of course, be occurring on the primary side or on the secondary side. Generally speaking, it seems that the overloading of one phase to the extent where 1-phase voltage would be appreciably reduced would be a temporary or random phenomena, and not a continued condition. In so far as this unbalance is a random phenomena, it will average out to be zero in the long run, and no customer has anything to be concerned about. If it is known in a particular installation that unbalance is a permanent condition and if the unbalance is actually large enough to affect the registration of the meters, some correction should be calculated and applied to the meter readings. It seems, however, that this will rarely be the case, and if it should be discovered to be the case, a more efficient alternate procedure would be to redistribute the loads on the lines so that they are more nearly balanced and so that the voltage is substantially balanced. The same thing can be said to be true of the 3-wire circuit; that is, if there is a permanent known unbalance in loading, the registration can be corrected for the error or the loads can be redistributed.

However, as a general conclusion, the only thing that can be said with certainty is that unbalances are random phenomena and, in the long run, are equally likely to be on any line, and the effect of one customer's unbalance on another customer averages out to be zero.

DINA, a Digital Analyzer for Laplace, Poisson, Diffusion, and Wave Equations

CORNELIUS LEONDES
STUDENT MEMBER AIEE

MORRIS RUBINOFF
ASSOCIATE MEMBER AIEE

IN AN extremely large number of problems concerning the behavior of physical systems the associated equation consists of the Laplace equation

$$\nabla^2 u=0 \qquad (1)$$

or one of its close relatives

Poisson equation $\nabla^2 u=\rho$ (2)

Diffusion equation $\nabla^2 u=k\dfrac{\partial u}{\partial t}$ (3)

Wave equation $\nabla^2 u=\dfrac{1}{v^2}\dfrac{\partial^2 u}{\partial t^2}$ (4)

For a given problem, equations 1 through 4 are not definitive as they stand, but require the specification of boundary and initial conditions peculiar to the problem at hand.

The importance of these equations is well known. Because these equations apply to so many physical phenomena, the need for their solution occurs with great frequency. Unfortunately, in most cases it is not possible to obtain the solutions to these equations through analytical procedures. Instead, resort must be made to special instruments or machines of one sort or another. The large number and the great variety of such instruments and machines constructed solely for the solution of these equations testifies to the importance of such devices.

The section of the references entitled "Devices" lists 27 papers describing both analogue and digital devices. This listing is merely suggestive, and is by no means meant to be comprehensive.

For simplicity of exposition, let us concentrate on the Laplace equation in two dimensions. This equation, generally expressed compactly in the notation of equation 1, may be written in many forms depending on the co-ordinate system chosen. Where the equation is solved analytically, the co-ordinate system is selected to ameliorate the difficulties of introducing boundary conditions. For example, if the boundary conditions are given on a circle, the natural choice is

$$\nabla^2 u=\frac{1}{r}\frac{\partial}{\partial r}\left(r\,\frac{\partial u}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}=0$$

On the other hand, when the Laplace equation is to be solved by numerical methods, the most convenient form of the equation appears to be the simple cartesian form

$$\nabla^2 u=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0$$

In this case, regardless of the shape of the bounding curve, the associated finite difference equation becomes particularly simple and from the standpoint of a computing machine, particularly easy to mechanize. It is, in fact, the purpose of this paper to describe the logical basis and general form of a relatively simple machine the *DINA* which is capable of solving the Laplace equation in cartesian coordinates (still with no restriction on the boundary conditions). In fact, as we shall see, the machine will be capable of solving for u even when u is meromorphic or when the related equations 1, 2, 3 or 4 are to be solved.

As previously mentioned, the first part of the references includes papers describing devices capable of analogue solution of these equations in two dimensions. Such devices include electrolytic tanks, rubber membranes, resistance networks, resistance-capacitance networks, resistance-inductance-capacitance networks, the differential analyzer, photoelastic pattern devices, fluid mapping devices, devices, employing optical methods, and so forth. These devices owe their success to the fact that the physical behavior of electrostatic fields, thin films in uniform tension, fluid flow, and so forth, are contained in equations 1 through 4. Conversely, any one device capable of solving these equations may be applied to solving problems of electrostatic fields, thin films, fluid flow, and so forth.

All the devices mentioned so far have analogue and are, therefore, of limited precision. It is well known that digital techniques are inherently capable of greater precision than analogue techniques. Moreover, some of these analogue devices are restricted in their capabilities. For example, the differential analyzer is not readily adapted to the solution of equations of the elliptic type, such as the Laplace equation with boundary values defined on a closed curve surrounding the region of interest. The machine proposed in this paper is of the digital type with consequent inherent higher precision.

The proposed machine is a special-purpose calculator. As a result, it has certain advantages over existing large-scale general-purpose digital computers in that it is designed specifically for the purpose at hand. In the first place, the DINA is to be so constructed as to yield maximum utilization of all its components, approaching 100-per-cent duty factor in its arithmetic unit. Second, the arithmetic unit is of a singularly elementary type, being devoid of multipliers or dividers and containing essentially only several adders. Third, it uses existing electronic digital computer techniques and in such a way as to take fullest advantage of their inherent speed.

Mathematical Basis of the DINA

The purpose of this section is to set down the equations used by the proposed computer in the numerical solution of equations 1 through 4. No attempt is made to develop these equations nor their inherent curtailment error. Nor is it possible in the restricted space of this paper to set down the methods already developed whereby the DINA may be used for more complex problems such as those involving time-varying boundary conditions, boundary conditions on multiply-connected regions, and so forth.

The numerical solution of partial differential equations involves the replacement of the analytic equation by a set of equivalent finite differences based on selected points within the region of interest. This method of analysis of physical problems actually predates the use of partial differential equations. In this method the difference equations are set up for the region and are solved by an iterative procedure. While the network of points may be distributed throughout the region in any prescribed manner, the use of a square (or rectangular) net generally provides the most satisfactory form of efficient numerical computation.

The difference equation approach to solving partial differential equations has received the attention of a great number of investigators over a long period of time. The references listed under "Mathematical Techniques" are suggestive of the

---

Paper 52-181, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted February 4, 1952; made available for printing May 19, 1952.

Cornelius Leondes and Morris Rubinoff are both of the University of Pennsylvania, Philadelphia, Pa.

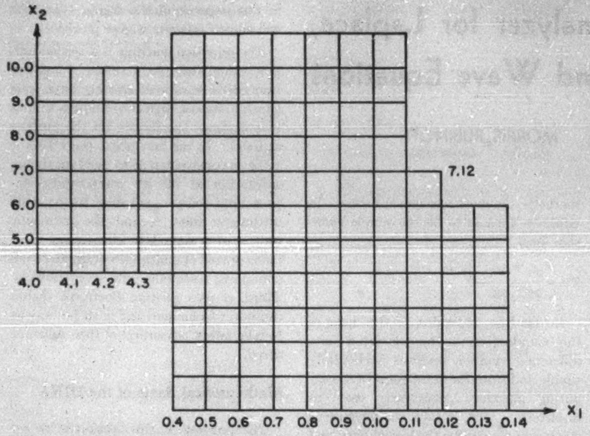


Figure 1. Network representation of an illustrative region of interest

amount of effort expended in this direction.

Consider first the difference equations for the n-dimensional Laplace equation, based on a square lattice with interval $h=\Delta x_i$. Then it can be shown that for the interior point $x=(x_1, x_2, x_3, \ldots x_n)$

$$u^{(m+1)}(x)=\sum_{j=1}^{n}\frac{u^{(s)}(x_j+h)+u^{(t)}(x_j-h)}{2n}$$

$$\underline{\underline{\Delta}}\ \frac{\Sigma u^{(k)}}{2n} \qquad (5)$$

where

$u^{(o)}(x)$ = initial guess of the value of u at x
$u^{(k)}(x)$ = the value of $u(x)$ after k applications of equation 5
$u^{(k)}(x_j\pm h)\equiv u^{(k)}(x_1, x_2, \ldots, x_j\pm h, \ldots, x_n)$

and $\underline{\underline{\Delta}}$ means "equal by definition."

For example, for $n=1$ and $n=2$ equation 5 becomes respectively

$$u^{(m+1)}(x)=\frac{u^{(s)}(x_1+h)+u^{(t)}(x_1-h)}{2} \qquad (6)$$

$$u^{(m-1)}(x_1, x_2)=[u^{(s)}(x_1+h, x_2)+ u^{(t)}(x_1, x_2+h)+u^{(v)}(x_1-h, x_2)+ u^{(w)}(x_1, x_2-h)]^{1/4} \qquad (7)$$

Note that these equations require for their computation only: 1. addition and 2. division by a number of the form 2^r, where r is an integer. If the numbers be displayed in binary notation, the division is replaced by a columnar shift r places to the right, an operation which is readily and economically mechanized.

In the 3-dimensional case, the denominator is six, which is not of the form 2^r. However, if desired, a fictitious fourth dimension x_4 may be invoked along with the assumption of symmetry in this direction. Then equation 1 becomes

$$u^{(m+1)}(x)=\sum_{j=1}^{4}\frac{[u^{(s)}(x_j+h)+u^{(t)}(x_j-h)]}{8}$$

$$=\sum_{j=1}^{3}\frac{[u^{(s)}(x_j+h)- u^{(t)}(x_j-h)+2u^{(m)}(x)]}{8} \qquad (8)$$

and again division may be eliminated.

To illustrate some possible procedures, consider the 2-dimensional case and the following two specific procedures:

The Liebmann method. Assume that the function $u(x, y)$ is required over the area shown in Figure 1 with the values of u given at all points on the boundary. The region is covered (implicitly) with a network of squares as shown in the figure, and the intersections are labelled according to their Cartesian co-ordinates. It is required to evaluate $u(x, y)$ for all interior points.

The Liebmann method consists of the systematic application of equation 7 with $s=t=m+1$, $v=w=m$. That is, the values of $u(x, y)$ in each column are improved one at a time from the bottom interior point up, from left to right; and at each application of equation 7, the most up-to-date values of $u(x, y)$ are utilized.

Since $u^{(m)}(x_0)$ is obsolete as soon as $u^{(m+1)}(x_0)$ has been computed, subsidiary storage is not required in the Liebmann method. In reference 37 it has been shown that the time required to reduce all initial errors by a factor of 10^6 is of the order of $5.6\,\tau N^4$, where τ is the time to perform one operation (division or addition) and N^2 is the number of small squares into which the (square) region has been divided.

The Richardson method. This method differs from the Liebmann method only in its use of the unimproved values of all the terms in $\Sigma u^{(k)}$. That is in applying equation 7, $s=t=v=w=m$. This method has two disadvantages. First, since in computing column c the unimproved values in column c-1 are required even though the improved values are available, at least one additional column of information must be stored. In practice, two auxiliary columns may be needed. Second, reference 37 shows that the computing time is doubled, about $11\tau N^2$ being required. On the other hand, it is now to be shown that the Richardson method is virtually identical with the method for solving the diffusion equation and hence may have merit in simplifying the structure of the proposed computer.

The diffusion equation is normally encountered in the form

$$\nabla^2 u=a^2\frac{\partial u}{\partial t} \qquad (9)$$

The corresponding finite difference equation is

$$\frac{\Sigma u^{(m)}-2nu^{(m)}(x)}{h^2}=a^2\,\frac{u^{(m+1)}(x)-u^{(m)}(x)}{t} \qquad (10)$$

where the notation is the same as before, but the superscripts m refer to the time co-ordinate $t=t_0+m\Delta t$

t_0 = time—zero
Δt = time increment

Thus, in 2-dimensions, we must choose $s=t=v=w=m$ (equation 7). In general, by choosing

$$\Delta t=h^2a^2/2n \qquad (11)$$

equation 10 becomes

$$u^{(m+1)}(x)=\frac{\Sigma u^{(m)}}{2n} \qquad (12)$$

which in appearance is identical with equation 5. Clearly, the Richardson method applies unchanged and consequently the solution is stable. It does not follow immediately that the solution of equation 12 is identical with the solution of equation 9 because differences of higher order have been neglected. However, a solution with accuracy sufficient for engineering purposes should result by choice of judiciously small h.

It should be observed that equation 12 differs from equation 5 in that at time t_0

the values of u are specified for all points of the region, both boundary and interior.

It is interesting to note also that because of the intimate connection between the Laplace equation and the diffusion equation, it should be possible to ascertain in advance the number of iterations required to solve equation 1 by the Richardson method from physical knowledge of the rate of diffusion for the associated pseudo diffusion equation.

The Poisson equation differs from the Laplace equation only in the presence of the term $\rho(x_j)$ in place of zero on the right-hand side of the equation. It is shown easily that the finite difference equation becomes

$$u^{(m+1)}(x)=\frac{\Sigma u^{(k)}-h^2\rho(x)}{2n} \tag{13}$$

Thus, both Liebmann's and Richardson's method may be applied with a simple modification consisting of one more addition operation. However, a complete table of $h^2\rho$, consisting of N^2 entries, must be provided in this case. Note that in the iteration process the evaluation of u at the point x_0 is accompanied by an inspection of $h^2\rho$ at the corresponding point x_0 in the $h^2\rho$ table, that is, the sampling of the two tables proceeds in synchronism.

Finally, the difference equation for wave propagation may be shown to be

$$u^{(m+1)}(x)=\frac{\Sigma u^{(m)}-u^{(m-1)}(x)}{n} \tag{14}$$

where $h^2=nc^2(\Delta t)^2$.

This is similar to the Poisson equation except that the $h^2\rho$ table is replaced by the function u at time $(m-1)$ and the denominator is n in place of $2n$. The remarks on equation 12 therefore apply here, with appropriate modification. The loss of the factor 2 is not a serious matter insofar as the mechanization of DINA is concerned.

For convenience, equations 5 through 14 are here collected together:

Laplace $\quad u^{(m+1)}(x)=\frac{1}{2n}\,\Sigma u^{(k)}$

Diffusion $\quad u^{(m+1)}(x)=\frac{1}{2n}\,\Sigma u^{(m)}$

Poisson $\quad u^{(m+1)}(x)=\frac{1}{2n}\,\Sigma u^{(k)}-h^2\rho(x)$

Wave propagation

$$u^{(m+1)}(x)=\frac{1}{n}\,\Sigma u^{(m)}-u^{(m-1)}(x)$$

Mechanization of DINA

The broad field of application of the proposed machine has been indicated in the preceding sections. Nevertheless, DINA is inherently a special-purpose computer and its construction will be justified only if it possesses one or more advantages over a general-purpose machine, such as:

1. Smaller physical size.
2. Lower capital expenditure.
3. Greater inherent speed.
4. Greater simplicity and ease of operation.

It is the purpose of this section to show how DINA could be mechanized so as to enjoy at least advantages 1, 2, and 3, and, with the possible exception of the required decimal-binary conversion, advantage 4 as well.

As in the previous sections, the discussion will be limited to the 2-dimensional equations. It should be noted that the extension to three (or more) dimensions is a straightforward one. However, in this case either the machine must be provided with a larger memory unit or the space under consideration must be treated as a number of smaller subspaces, each with a cubic network of, say, 30 by 30 by 30 intervals.

An input unit would, of course, be required. The sole function of this unit would be to introduce into DINA the fixed boundary values for the elliptic equations, and the fixed and/or time-varying boundary values and initial values for the parabolic and hyperbolic equations. Note that external program instructions would not be required since the program sequence would be fixed once and for all in the initial construction of the machine. The minor modifications required for the different types of equations could be incorporated easily by the inclusion of a few switches on a control panel.

For the Laplace and diffusion equations, only one full 2-dimensional table is required. In order to satisfy requirements 1, 2, and possibly 4, this table would be stored serially and in binary notation on a magnetic drum (or disk) in the manner indicated in Figure 2. It is estimated that a table of 100 rows by 100 columns can be accommodated on a drum 6 inches in diameter and 4 inches long, the numbers being represented to a precision of 1 part in 10^6. An auxiliary table of two columns would be required for interim values; this does not increase the length of the drum appreciably.

For equations 13 and 14, a second full table is required to store $\rho(x, y)$ and $u^{(m-1)}$ respectively. For the former, the auxiliary table would be required as well; for simplicity of programming, it would be employed in the latter case as well. All three tables can be accommodated if the Laplace drum be increased to 8 inches in length, without change in diameter.

A number of channels on the drum would be reserved for clock-pulse and control-pulse generation, requiring at most 1 inch additional drum length. Thus the drum would not exceed 9 inches length by 6 inches diameter, which is compact in size and capable of rotation at upwards of 100 revolutions per second. (In this regard it might be mentioned that a slightly larger magnetic drum has been successfully operated at 200 revolutions per second.)

The amount of equipment that would be required to mechanize the arithmetic unit is best visualized through reference to the flow diagram shown in Figure 3. Consider first the solution of the Laplace (or diffusion) equation. To start the problem, boundary values (and initial values) are inserted into Table I through the distribution pyramid P_1, while Table II is either erased to zero or disconnected electronically from adder A_2. The input is then disconnected (except when the boundary conditions are time-varying)

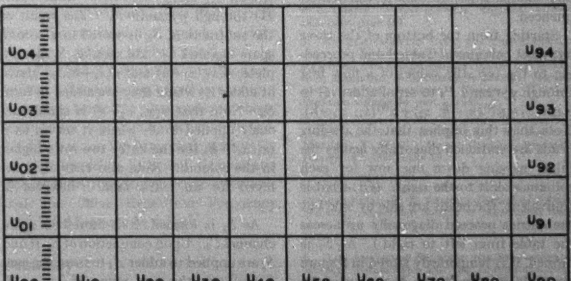


Figure 2. Method of information storage on a magnetic drum

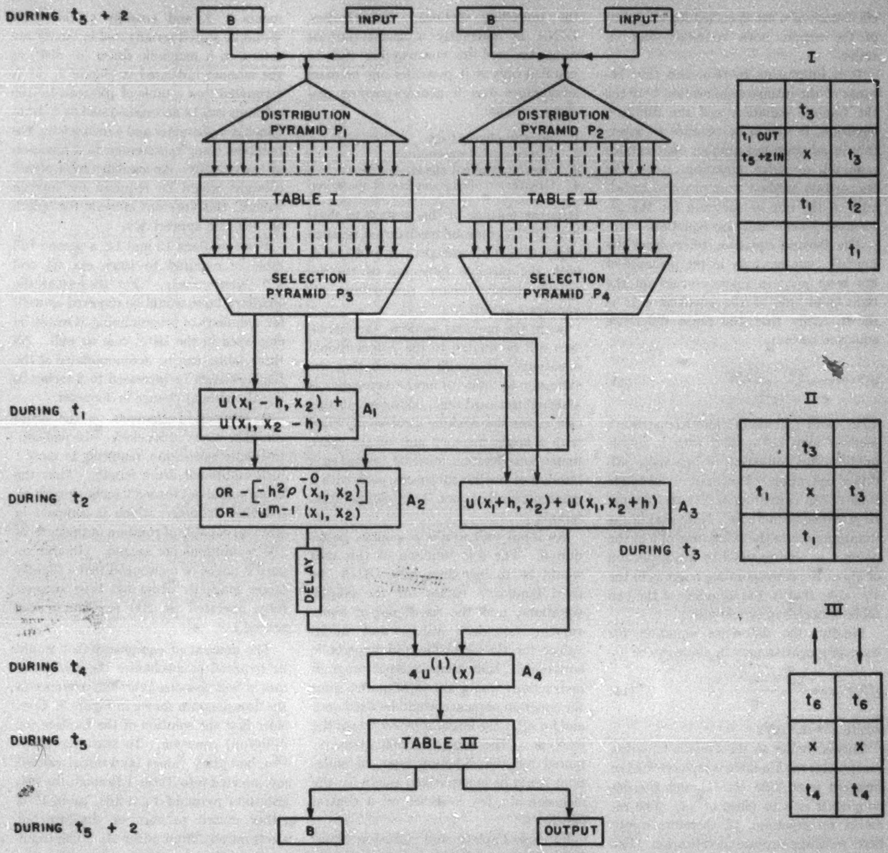

DURING t5 + 2
B
INPUT
B
INPUT
DISTRIBUTION PYRAMID P1
DISTRIBUTION PYRAMID P2
TABLE I
TABLE II
SELECTION PYRAMID P3
SELECTION PYRAMID P4
DURING t1
u(x1 - h, x2) + u(x1, x2 - h)
A1
DURING t2
-0
OR - [-h²ρ (x1, x2)]
OR - u^(m-1) (x1, x2)
A2
u(x1+h, x2) + u(x1, x2+h)
A3
DURING t3
DELAY
DURING t4
4u^(1)(x)
A4
DURING t5
TABLE III
DURING t5 + 2
B
OUTPUT
I
t3
t1 OUT
t5+2IN
x
t3
t1
t2
t1
II
t3
t1
x
t3
t1
III
t6
t6
x
x
t4
t4

Figure 3. Block diagram of DINA

and the machine initiated to run self-sequenced.

Starting from the bottom of the three left-most columns of Table I and proceeding to the top, the values of u flow first through pyramid P_3 to serial adder A_1 to form $S_1=u^{(0)}(x_1-h,\ x_2)+u^{(0)}(x_1,\ x_2-\mathrm{h})$. Note that this implies that the reading heads are stationed diagonally across the table, moving down one row for each columnar shift to the right. (Or, what is equivalent, the heads are side by side but the entries proceed diagonally up across the table from left to right.) As S_1 is formed, it is temporarily stored in a spare channel C_1 on the drum. At the end of the number interval t_1, when S_1 has just been completely computed, S_1 is fed into subtractor A_2 along with zero (from Table II) through pyramid P_4. The result of the subtraction, S_2, is stored in a second spare channel C_2. As soon as S_2 is complete, $u(x_1,\ x_2+h)$ and $u(x_1+h,\ x_2)$ arrive at adder A_3 where they are added to form S_3. Note that $u(x_1,\ x_2+h)$ is simultaneously applied to A_1 where it serves as $u(x_1,\ x_2'-h)$ for the entry two rows higher in the column. Note also that S_2 is delayed for an extra number-interval in C_2.

As S_3 is formed, it is stored in spare channel C_3. Upon completion of S_3, it and S_2 are applied to adder A_4 to form the sum $S_4=u^{(1)}(x)$, the new value of $u(x)$. S_4 is stored in one of the two columns of Table III.

The adders A_1 through A_4 are single-binary-digit adders, each of which requires only about 20 crystal diodes and three pentodes. Clearly then, the selection pyramids associated with the three Tables, particularly those for Tables I and II, account for almost all the electronic equipment in the machine.

Note that the process is a continuous one, so that all the adders are operating full time, that is, the whole arithmetic unit operates all the time, giving 100-per-cent utilization of components. While numbers are being entered into one column of Table III, numbers are removed from the other column, but two pulse times later than usual (that is, divided by 4); these are inserted into Table I. A study of the process shows that new values of $u(x)$ are reinserted into their correct locations in Table I four number-intervals after the old u-value is no longer needed. Note that Table III must be controlled so that the roles of the two columns are interchanged at the completion of each columnar computation.

The process continues from left to right to the end of the table, then jumps discontinuously back to the left-most columns and repeats indefinitely. With the drum rotating at 100 revolutions per second, the Laplace equation is solved to an accuracy of 1 per cent of the initial error in about 3 hours or to one part in 10^6 in about 8 hours. Comparing this iteration speed of DINA for a 100 by 100 lattice with the iteration speed of UNIVAC for the Laplace boundary value problem[40] it follows that DINA is over 100 times faster for problems of this order of magnitude.

In the case of the diffusion equation, a complete table of values of $u^{(m)}(x)$ is read out of the machine at prescribed times, read-out being most easily mechanized from Table III at point B. Also, if the boundary conditions are time-varying, they must be read into Table I from the input unit when required. This demands additional control channels.

For the Poisson equation, the process is modified only slightly. In this case, Table II is inscribed with the values of $-h^2\rho(x,y)$, and the subtractor A_2 forms $S_2 = S_1 - (-\mathrm{h}^2\rho) = S_1 + \mathrm{h}^2\rho$. It follows from equation 13 that no further modification need be made.

In the solution of the wave equation, the flow is automatically such that A_2 forms $S_2 = S_1 - u^{(-1)}(x)$, as required. However, it is now necessary that the roles of Tables I and II be interchanged at the completion of each computation over the whole table. For that reason, an appropriate control pulse must be generated from a control channel on the drum; this pulse might be the same as that which steps the column selection from the extreme right to the extreme left. At the same time, it is necessary to interchange the roles of P_1 and P_2, and of P_3 and P_4. For this reason, three outlets are provided from P_4, and access into P_2 from B. Finally, it is required that the read-out from B be only 1 pulse-time late. This may be achieved by generating the here unwanted pulse-time delay electronically and providing a switch on the control panel to by-pass the delay in this case.

It is vital to the operation of DINA that the machine be capable of distinguishing between entries which are boundary conditions and those which are not. For this purpose an extra pulse position is provided preceding each number-location. A boundary value will be recognized by the presence of a pulse in this position; this pulse then serves to inhibit transfer from Table III into the corresponding number-location in Table I. The provision of a dead space between numbers wherein subsiding control pulses may be inserted might prove useful in other connections; the possibilities of such a scheme are too lengthy for inclusion.

Summarizing the preceding remarks, it is observed that DINA requires the following units:

1. An input unit.
2. Two full tables and one 2-column table of number storage.
3. Four gating pyramids.
4. Four serial adders.
5. Four temporary-storage channels.
6. Several control channels, including a clock-pulse generating channel.
7. An output unit.

A few words to justify the preference of magnetic storage are in order. Certainly, of all existing memory units, it provides maximum storage density and consequently minimum size. In this regard, it should be recalled that a competent DINA should store more than 20,000 numbers, or over 400,000 binary digits. The cost of storing so many digits by other existing methods would certainly be prohibitive for a special-purpose machine, or, for that matter, for a general-purpose one.

Again, the permanence of magnetic storage is ideal for indefinite retention of control information and clockpulses, which are invariant in DINA.

Third, the major objection to magnetic drum storage is its long access time. But access time is of no consequence in the process previously outlined since the numbers are automatically positioned to be read out without a waiting period. From the standpoint of speed, the only objection which might justifiably be raised is that of basic pulse repetition frequency. Computation of pulse repetition frequency from the assumptions above leads to an approximate value of 250 kc. This is not too far from that achievable by mercury delay lines and electrostatic storage tubes; moreover, it is conceivable that the Pulse Repetition Frequency postulated here can be raised.

Summary

It has been the purpose of this paper to present the logical basis of a new type of digital computer. This computer has been contrived specifically for the solution of a class of partial differential equations. Because of the nature of the process utilized in these solutions, the name DINA (Digital Network Analyzer) has been found to be most appropriate for this machine.

Since this computer has been devised specifically for the purpose of solving partial differential equations the resulting physical size of the machine has been shown to be relatively small. Furthermore, maximum utilization of the circuits has resulted. Of existing large-capacity storage devices, the magnetic drum is suggested as having every advantageous characteristic with the exception of speed.

When compared with analogue computers which are capable of handling the same class of problems, this machine enjoys the advantages of greater versatility and higher precision of number representation. When compared with large-scale electronic digital computers, the DINA enjoys the feature of maximum utilization of circuits and is thereby considerably faster and smaller. In addition, it is not necessary to program problems for solution on the DINA since by the nature of the machine this programming is already built into its logic.

As a final remark it should be pointed out that in the future it seems almost certain that the principles on which DINA is based will be capable of extension to the handling of a broader class of problems.

References

Devices

1. Construction of a Complete Set of Solutions of a Linear Partial Differential Equation in Two Variables, by Use of Punch Card Machines, **Stefan Bergman**. *Quarterly of Applied Mathematics*, Brown University (Providence, R. I.), volume 4, 1946–47, page 233.

2. Analogue Calculating Machine for Functions of a Complex Variable, **A. R. Boothroyd, E. C. Cherry.** *Nature* (London, England), volume 163, 1949, page 687.

3. An Electrolytic Tank for the Measurement of Steady State Response, Transient Response and Allied Properties of Networks, **A. R. Boothroyd, E. C. Cherry, R. Makar.** *Journal*, Institution of Electrical Engineers (London, England) volume 96, part 1, 1949.

4. Electrolytic Field Plotting Through for Circularly Symmetric Systems, **M. Bowman-Manifold, F. Nicoll.** *Nature* (London, England), volume 140, 1938, page 39.

5. Numerical and Network Analyzer Solutions of the Equivalent Circuits for the Elastic Field, **G. K. Carter.** *Journal of Applied Mechanics*, American Society of Mechanical Engineers (New York, N. Y.), volume 11, 1944, page 162.

6. The Analogies Between the Vibrations of Elastic Membranes and the Electromagnetic Fields in Guides and Cavities, **E. C. Cherry.** *Journal*, Institution of Electrical Engineers (London, England), volume 96, part 3, 1949, page 346.

7. Determination of Aperture Parameters

By Electrolytic Tank Measurements, **S. B. Cohn.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 39, 1951, page 1416.

8. A Resistor Network for the Solution of the Laplace Equation, **D. C. De Packh.** *Review of Scientific Instruments*, American Institute of Physics (New York, N. Y.), volume 18, 1947, page 798.

9. Punched Card Methods in Scientific Computation (book), **W. J. Eckert.** Columbia University Press, New York, N. Y., 1940.

10. Mechanical Tracer for Electron Trajectories, **D. Gabor.** *Nature* (London, England), volume 139, 1937, page 373.

11. Experimental Determination of Impedance Functions by the Use of an Electrolytic Tank, **W. W. Hansen, O. C. Lundstron.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 33, 1945, page 528.

12. Differential Analyser, **R. D. Hartree.** *Monograph Number 17-502*, Permanent Records of Research and Development, Ministry of Supply (London, England).

13. A Method for the Numerical or Mechanical Solution of Certain Types of Partial Differential Equations, **R. D. Hartree, J. R. Womersley.** *Proceedings*, Royal Society of London (London, England), volume 161, 1937, page 353.

14. Calculating Instruments and Machines (book), **R. D. Hartree.** University of Illinois Press (Chicago, Ill.), 1949, page 26.

15. Factors Influencing the Design of a Rubber Model with Demonstrations of Other Methods of Solving the Laplace Equation. *Journal*, Institution of Electrical Engineers (London, England), volume 97, part 2, 1950, page 439.

16. The Equivalent Circuits of the Elastic Field, **G. Kron.** *Journal of Applied Mechanics*, American Society of Mechanical Engineers (New York, N. Y.), volume 11, 1944, page 149.

17. Equivalent Circuit of the Field Equations of Maxwell, **G. Kron.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 32, 1944, page 289.

18. Numerical Solution of Ordinary and Partial Differential Equations by Means of Equivalent Circuits, **G. Kron.** volume 16, 1945, page 172.

19. Electric Circuit Models of Partial Differential Equations, **G. Kron.** *Electrical Engineering*, volume 67, July 1948, pages 672–84.

20. An Electrolytic Tank for Exploring Potential Field Distributions, **R. Makar, A. R. Boothroyd, E. C. Cherry.** *Nature* (London, England), volume 161, 1948, page 846.

21. Fields from Fluid Flow Mappers, **A. D. Moore.** *Journal of Applied Physics* (New York, N. Y.), volume 29, 1950, page 790.

22. The Determination of Temperature Transients in Cable Systems by Means of an Analogue Computer, **J. H. Neher.** *AIEE Transactions*, volume 70, part II, 1951, pages 1361–71.

23. **V. Paschkis, H. D. Baker.** *Transactions*, American Society of Mechanical Engineers (New York, N. Y.), volume 64, 1942, page 102.

24. Electrical Network Analyzers for the Solution of Electromagnetic Field Problems, Parts I and II, **K. Spangenberg, G. Walters, F. Schott.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 37, 1949, pages 724 and 866.

25. The Solution of Torsion Problems by Means of a Liquid Surface, **C. Sunatani, T. Matuyama.** *Technology Reports*, Tohoku Imperial University (Japan), volume 12, 1938, page 374.

26. Numerical Solution of the Boundary Value Problem for the Potential Equation by Means of Punched Cards, **M. Tormes.** *Review of Scientific Instruments*, American Institute of Physics (New York, N. Y.), volume 14, 1943, page 248.

27. Factors Influencing the Design of a Rubber Model, **G. B. Walker.** *Journal*, Institution of Electrical Engineering (London, England), volume 29, part 2, 1949, page 319.

Mathematical Techniques

28. The Approximation of Functions Satisfying a Linear Partial Differential Equation, **Stefan Bergman.** *Duke Mathematical Journal*, Duke University (Durham, N. C.), volume 6, 1940, page 537.

29. Determination of the Axial Potential Distribution in Axially Symmetric Electrostatic Fields, **S. Bertram.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 28, 1940, page 418.

30. Method of Finite Differences for the Solution of Partial Differential Equations, **R. Courant.** *Annals*, Computation Laboratory of Harvard University (Cambridge, Mass.), volume 16, 1948, page 153.

31. A Practical Method for the Numerical Evaluation of Solutions of Partial Differential Equations of the Heat Conduction Type, **J. Crank, P. Nicolson.** *Proceedings*, Cambridge Philosophical Society (London, England), volume 43, 1947, page 50.

32. The Potential Analogue Method of Network Synthesis, **S. Darlington.** *Bell System Technical Journal* (New York, N.Y.), volume 30, 1951, page 315.

33. Stability in the Numerical Solution of Initial Value Problems in Partial Differential Equations, **R. P. Eddy.** *Naval Ordnance Laboratory Memorandum 10232* (Washington, D. C.), 1949.

34. Numerical Methods of Solving Partial Differential Equations, **H. W. Emmons.** *Quarterly of Applied Mathematics*, Brown University (Providence, R. I.), October 1944.

35. Analysis of Numerical Solutions of Transient Heat-Flow Problems, **C. M. Fowler.** *Quarterly of Applied Mathematics*, Brown University (Providence, R. I.), volume 3, 1946–47, page 361.

36. A Short Account of Relaxation Methods, **L. Fox.** *Quarterly Journal of Mechanics and Applied Mathematics* (London, England), volume I, September 1948, pages 253–80.

37. Convergence Rates of Iterative Treatments of Partial Differential Equations, **S. Frankel.** *M.T.A.C.*, volume 4, 1950, page 65.

38. A Method for the Numerical or Mechanical Solution of Certain Types of Partial Differential Equations, **D. Hartree, J. Womersley.** *Proceedings*, Royal Society of London (London, England), series A, volume 161, 1937, page 353.

39. On the Precision of a Certain Procedure of Numerical Integration, **H. Huskey.** *Journal of Research*, National Bureau of Standards (Washington, D. C.), volume 42, 1949, page 57.

40. Mathematical Tables and Other Aids to Computation (book). "Coding of a Laplace Boundary Value Problem for the UNIVAC," **Hubert Livingston, Frances Snyder,** volume 3, 1949, page 341.

41. The Numerical Solution of Laplace's and Poisson's Equations, **D. Moskovitz.** *Quarterly of Applied Mathematics*, Brown University (Providence, R. I.), volume 2, 1944, page 148.

42. Calculation of the Electromagnetic Field Frequency, and Circuit Parameters of High Frequency Resonator Cavities, **H. Motz.** *Journal*, Institution of Electrical Engineers (London, England), volume 93, part III, 1946, page 335.

43. A Study of the Numerical Solution of Partial Differential Equations, **G. O'Brien, M. Hyman, S. Kaplan.** *Journal of Mathematics and Physics*, Massachusetts Institute of Technology (Cambridge, Mass.), volume 29, 1950, page 223.

44. The Approximate Arithmetical Solution by Finite Differences of Physical Problems Involving Differential Equations, **L. F. Richardson.** *Philosophical Transactions*, Royal Society of London (London, England), series A, volume 210, 1910, page 307.

45. The Numerical Solution of Laplace's Equation, **G. Shortley, R. Weller.** *Journal of Applied Physics* (New York, N. Y.), volume 9, 1938, page 334.

46. Relaxation Methods in Engineering Science (book), **R. V. Southwell.** Oxford University Press, London, England, 1940.

47. On a Method of Approximate Solution of a Plane Harmonic Function, **C. Suratani, S. Negoro.** *Technology Reports*, Tohoku Imperial University (Japan), volume 12, 1938, page 339.

48. Stability of Solution of Partial Differential Equations, **L. Thomas.** *Naval Ordnance Laboratory Report 1132*, 1950, page 83.

49. Numerical Methods in the Theory of Vibrating Bodies, **A. Vaysonyi.** *Journal of Applied Physics* (New York, N. Y.), volume 15, 1944, page 598.

50. Improvements to the Accuracy of Arithmetical Solutions to Certain Two-Dimensional Field Problems, **L. Woods.** *Quarterly Journal of Mechanics and Applied Mathematics* (London, England), volume 3, 1950, page 349.

Discussion

J. R. Whinnery (Hughes Aircraft Company, Culver City, Calif.): The special-purpose digital machine is to be compared with special purpose analogue computers on the one hand and with general-purpose digital computers on the other. The authors have made this comparison, pointing out that precision is the pertinent factor in the first comparison, and simplicity, lower cost, and increased speed the advantages in the second. The choice for a given laboratory obviously will depend upon the number of problems of the special class to be solved and the accuracy required and there seems little doubt that the machine described would find its place, though in some cases the analogue or general-purpose digital machines would be preferable.

Elaborating a bit, the electrolytic tank has been found to be an indispensable device, not only for electrostatic problems such as gun design but also as an analogue computer for networks, antenna arrays,[2] and other functions of a complex variable.[3] Incorporating improvements such as those described by Boothroyd, Cherry, and Makar,[4] it can be made to give very good accuracy. There is no tool of comparable simplicity and general usefulness for the wave equation, which is the equation we have most occasion to use in microwave work. It would be possible to build an analogue analyzer with lumped circuit elements which might be connected to solve any of the equations treated here (Laplace, Poisson, diffusion, or wave equation) and others, using the ingenious equivalent circuits devised by Kron.[5] About all that has been done in this direction has been the use of existing network analyzers built for power system studies,[6] and the construction of analyzers for Maxwell's equations with plane or cylindrical symmetry.[7] It may be possible from these efforts to make a comparison of size and cost between DINA and a network analyzer general enough to solve all of the same problems. If DINA is simpler, it would be chosen almost certainly on the basis of the greater possible accuracy, though there are certain problems where the analogue approach is ideal in that it fits the experimental procedures of the physical problems to be solved. An example is that in which a single value of electric field across the gap of a resonator is desired as conducting boundaries are varied in shape. This single value can be measured on the network analyzer as boundary conditions are changed, whereas the numerical solution requires that each problem be solved completely.

In comparing DINA with the general-purpose digital machines, not all of the stress should be placed on smaller size and decreased cost, for the increased speed mentioned as an advantage in the section "Mechanization of DINA" in the paper may make it possible to do problems for which most general-purpose machines are impractical. For example, 3-dimensional field problems are frequently too lengthy to justify computation on general-purpose computers. Much would be gained then if the special-purpose machine brought these into the range of practicality.

Finally, it is noted that straightforward difference methods are used in planning the machine. The authors might like to comment as to whether the relaxation methods described by Southwell[8,9] which save much time in desk computations where judgment can be used have any advantage in machine calculations.

REFERENCES

1. See reference 32 of the paper.

2. APPLICATIONS OF POTENTIAL THEORY TO THE DESIGN OF LINEAR ARRAYS, Taylor and Whinnery. *Journal of Applied Physics*, American Institute of Physics (New York, N. Y.), volume 22, January 1951, page 19.

3. See reference 2 of the paper.

4. See reference 3 of the paper.

5. See reference 16 of the paper.

6. See reference 5 of the paper.

7. See reference 24 of the paper.

8. See reference 46 of the paper.

9. RELAXATION METHODS IN THEORETICAL PHYSICS (book), R. V. Southwell. Oxford University Press, London, England.

Morris Rubinoff: As mentioned in the paper, a special-purpose machine is justified when its cost is kept low. In the case of DINA, this is a direct result of the use of the Richardson method of solution. Note that even the use of the Liebmann method for the Laplace equation complicates the machine for solving the other equations. Relaxation methods would require drastic modifications to DINA with large increase in cost.

It should be kept in mind that all digital computers, whether special- or general-purpose, are incapable of performing (useful) computations other than those previously programmed by a mathematician. Thus, no existing or currently planned digital computer is capable of utilizing the relaxation procedure, at least where "judgment" must be used.

The Variation at Constant Density of the Dielectric Breakdown of Paper with Air Resistance

PAUL CLOKE
FELLOW AIEE

KRISHNA K. KHANDELWAL
NONMEMBER AIEE

FOR many years engineers have been concerned to find what the factors are that bear on and control the dielectric strength of electric cable insulating papers. Among the well-known factors investigated is the paper's density which, in most instances, has been inextricably intertwined with the paper's air resistance. As far as the authors of this article are aware, no serious attempt has been made yet to separate the influence of air resistance as a single factor. It is the belief of the authors that they have succeeded in making this separation.

Some years ago work of the nature of this investigation was touched upon at the laboratories of the General Cable Corporation. However, the present investigation is the first comprehensive survey in which the density has been held constant and the air resistance varied thru rather wide limits. The density control was accomplished since it was possible to control also both the beating time and the calender pressure. The tests in general were facilitated by the fact that the University of Maine has the unusual combination of a well-equipped electrical engineering laboratory and a well-equipped pulp and paper laboratory.

This paper is the third in a series of which the first concerned a specially shaped electrode which obviated in very great degree the phenomenon of edge effect. The first tests conducted were made with an impulse wave.[1] The second paper reported the results of subsequent tests made on the same paper with the special electrode using a 60-cycle wave.[3] The paper used for these first two investigations was secured from paper manufacturers.

The curve of Figure 5 shows very clearly the relationship of dielectric breakdown to air resistance at constant density of 0.760 gram per cubic centimeter. The results of the experiments further show that dielectric strength increases with the increase of either density or air resistance; but the degree of control of dielectric strength with either property is limited.

Apparatus

THE SPECIAL ELECTRODE

A unique feature of all the tests was a specially shaped electrode whereby the breakdowns, being confined to a flat central part of the electrode, avoided in great degree the phenomenon of edge effect. Figures 1 and 2 give an idea of this special electrode. Figure 1 is, of course, not to scale, and Figure 2 shows merely

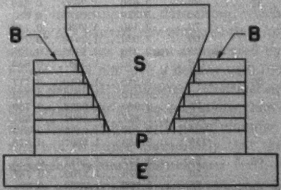


Figure 1. Special electrode and build-up paper (not to scale)

Paper 52-210, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted January 28, 1952; made available for printing May 13, 1952.

PAUL CLOKE is with the University of Maine, Orono, Maine. KRISHNA K. KHANDELWAL is with Syracuse University, East Syracuse, N. Y.

The authors wish to acknowledge the aid given by the staff of the Department of Chemical Engineering at the University of Maine. They are particularly grateful to R. W. Atkinson, Chief Research Engineer for the General Cable Corporation, for his many helpful and constructive comments and suggestions and for reviewing the manuscript; to Mr. R. G. Simonton of the Union Mills Paper Manufacturing Company, and to Mr. E. G. Ham of the John A. Manning Paper Company, Inc., who sent paper which was used in the first two investigations. The authors also feel especially indebted to Mr. C. F. Terrell, Instructor in English at the University of Maine, for reviewing the manuscript, making many helpful rearrangements, and rewriting several parts of the paper.

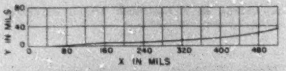


Figure 2. Plot of curved portion of electrode

the shape of the outer curved portion. The paper indicated in reference 1 gives more details concerning the electrode's design. One important assumption in its design is that the maximum stress transversely to the plane of the paper tested will not be more than 2 per cent of the radial stress in order to insure that failures will result from the radial stress and not from the tangential component. The transverse strength was assumed to be fifty times the strength parallel to the paper's surface.

The electrode *S* is flat for a central portion of 1/2-inch radius. *E* is the flat bottom electrode and *S* the special electrode. *E* is either: 1/2-mil or 1-mil rice paper built up to a total thickness of approximately 5 mils and cut to fit the upper special electrode; or a piece of the same paper as that tested, cut and sandpapered with very fine sandpaper to fit the upper special electrode. Paper prepared in this latter manner was most often used. All electrodes were kept highly polished.

SUPPLEMENTARY EQUIPMENT

A General Electric Company 15-kva 150,000-volt testing transformer was used to make the breakdowns. This transformer was supplied by a General Electric voltage regulator driven by a d-c motor through a system of gears. A filter was placed between the regulator and the transformer to iron out harmonics in the impressed wave. For paper making the following were used: A Williams pulp disintegrator, a Williams fiber disintegrator, and a Williams hydraulic press. An electric oven with automatic temperature control and a vacuum pump were used to treat the paper samples. In all 60-cycle researches the voltage was raised at the uniform rate of 500 volts per second. The point of breakdown was noted with a potential transformer and calibrated voltmeter. Sphere gaps and a high voltage voltmeter were used as additional means of checking the meter.

Table I

Set A Beating Time, Minutes	Pressure at Press, Pounds Per Square Inch	Average Thickness Per Sheet, Mils	Average Weight Per Sheet, Grams	Density, Grams Per Cubic Centimeter
0	25	4.680	1.1858	0.466
	50	4.040	1.2838	0.585
	75	3.347	1.2120	0.666
	100	3.300	1.2213	0.682
	125	3.113	1.2138	0.718
	150	3.060	1.2440	0.749
15	25	3.620	1.2600	0.640
	50	3.320	1.2423	0.689
	75	3.140	1.2685	0.743
	100	3.020	1.2940	0.788
	125	2.910	1.2405	0.785
	150	2.750	1.2537	0.839
30	25	3.240	1.2610	0.717
	50	3.013	1.2500	0.763
	75	2.960	1.2513	0.780
	100	2.827	1.2607	0.821
	125	2.700	1.2462	0.849
	150	2.653	1.2607	0.875
45	25	3.147	1.2647	0.738
	50	2.970	1.2700	0.787
	75	2.700	1.2588	0.858
	100	2.640	1.2408	0.865
	125	2.700	1.2893	0.878
	150	2.673	1.2678	0.873

Table II

Set	Average Thickness	Average Density, Grams Per Cubic Centimeter	Air Resistance, Seconds	Breakdown Strength Volts Per Mil 60 cycles	Mean Deviation
B	3.00	0.765	1,384.0	2,165	99
C	2.96	0.809	364.1	2,075	105
D	3.15	0.760	168.7	1,975	114
E	3.09	0.772	49.8	1,717	151

Procedures

PREPARATION OF PULP

The Technical Association of the Pulp and Paper Industry (TAPPI) standard procedures for beating,[3] sheet formation,[4] and sheet testing,[5] were followed in preparing the handsheets, except for changes of apparatus indicated under the section entitled "Supplementary Equipment."

Briefly, 360 grams of bone dry pulp were slurried in 23 liters of water and charged into the standard beater at a temperature of 23 ±2 degrees centigrade. A load of 5,500 grams was added to the lever arm of the bedplate and beating was started the moment the bedplate was loaded. Sufficient stock for the required number of sheets was withdrawn from the beater at the end of each desired beating time. The stock was diluted to a consistency of 0.15 per cent after being defibered in a Williams fiber disintegrator.

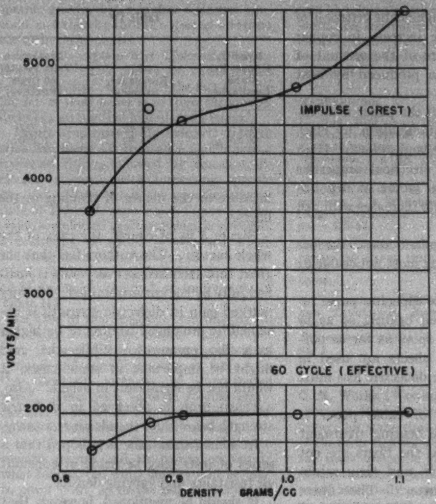
PREPARATION OF SHEETS

Sheets were made in a British handsheet machine and placed between mirror-polished plates covered with blotters in each case. When a stack of seven sheets had been made, they were placed in the Williams press and pressed for 5 minutes at the desired pressure. The Williams press was calibrated for a 64-square-inch surface instead of the TAPPI standard 200-square-centimeter surface. The pressures were recorded as read since the areas compared were the same. The sheets were then placed in the constant humidity room at constant temperature and relative humidity of 50 per cent. Readings of air resistance were made in a Gurley Densometer. Density and thickness were then taken with an analytic balance and micrometer. Sheets were prepared at pressures of 25, 50, 75, 100, 125, and 150 pounds per square inch for each of the four beating times of 0, 15, 30, and 45 minutes. The calculated results are shown in Table I and Figure 4.

In obtaining data for Table II and Figure 5, about 50 sheets were made for each beating time and at the pressure which the curves of Figure 4 indicated would give a density of 0.760 grams per cubic centimeter. These sheets were treated and measured in the same way as were those for Figure 4.

After the sheets were properly prepared as indicated they were placed under vacuum of 2 millimetres of Hg and at a temperature of 105 degrees centigrade for 24 hours. WEMCO "C" oil was then admitted. Each sample was then immediately tested, in a stone crock containing the same kind of oil, between the electrodes as previously described. The rate of application of the voltage was 500 volts per second. The breakdown voltage was measured easily and accurately simply by throwing a switch which started a d-c motor which drove the voltage regulator at the set and desired speed and by observing the point on the voltmeter at which the sample broke down, which point was sharp and very definite. Each sheet was tested for breakdown, one at a time, in at least three different places and the average taken. The breakdown voltage was calculated from the average thickness which was measured in four or five different places on each sheet with a calibrated micrometer and averaged to get the average thickness which in all cases was approximately 3 mils. The weight of each sheet was taken, after proper preparation as indicated elsewhere in the paper, by a highly accurate analytical balance. The oil in the crock was filtered after each test run.

Usually the oil used for impregnating the paper samples was heated at 105 de-



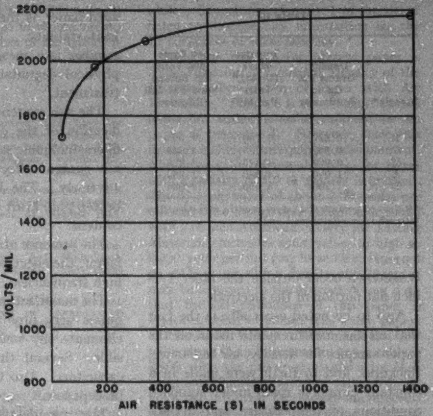


Figure 3 (left). Comparison of breakdown with impulse and with 60-cycle waves

Figure 5 (above). Breakdown versus air resistance

grees centigrade for 24 hours. However, repeated tests showed that this was unnecessary as the oil was free of moisture.

Discussion

The data in Table III and the plot of this data in Figure 3 were taken from the work done by Cloke and Bates[1] and by Whitten.[2] It was suggested, after the work was done on the impulse waves, that it would be well to repeat the tests on the same paper for 60 cycles. This was done in the investigation by Whitten,[2] and is here reported for the first time. The paper used in each case was furnished by paper manufacturers and was from the runs being made for cable work.

Two things stand out and are worthy of special note. In the first place the breakdown strength in volts per mil for the impulse waves, as seen from Table III and the curves of Figure 3 is much higher than for the same paper at 60 cycles. (Effective values are given for these latter and crest values for the impulse wave.) Furthermore, the general trend of the curves for 60 cycles in this investigation[2] and those of Figure 5 are very similar.

It should be noted also that there is a sharp rise in dielectric strength for the impulse waves at higher density, above 1.0 gram per cubic centimeter. This is doubtless of physical significance and is possibly tied in with greater compactness of paper fibers.

As the authors expected, a series of curves between density and pressure at each of the four beating times indicated that a constant density line would cross all the curves for the separate beating times. By inspecting the curves in Figure 4, one sees that a constant density line of 0.760 gram per cubic centimeter does cross all the pressure lines except for the curve for 0 beating time. For this curve, however, the pressure was obtained by minor extrapolation.

By inspecting the data from which Figure 5 is plotted, one observes that for some points the pressure as indicated by Figure 4 did not give a density of 0.760 gram per cubic centimeter. Considering the complexity of the operations, however, those familiar with paper manufacture will recognize immediately that the agreement is remarkably close.

The efficacy of the special electrode is attested by the fact that in only two of the 591 breakdown observations did the

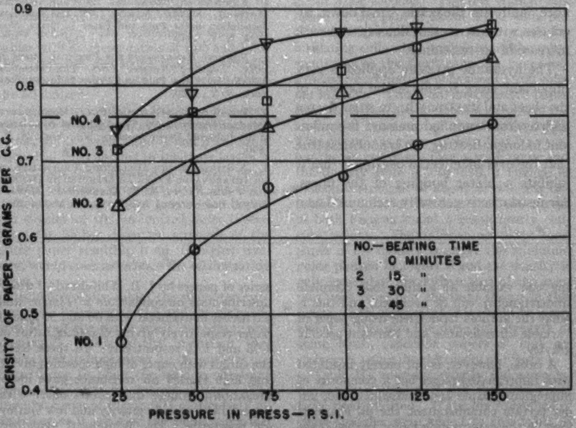


Figure 4. Density of paper versus pressure press

Table III

Sample*	Density, Grams Per Cubic Centimeter	Impulse Wave (Crest), Volts Per Mil	60 Cycle 500 Volts Per Second Volts Per Mil (Effective)
B	0.827	3,740	1,699
C	0.880	4,635	1,926
E	0.909	4,545	1,928
A	1.010	4,850	1,985
D	1.106	5,500	2,035

*Refers to commercial paper used in investigations referred to in references 1 and 2.

breakdown occur outside the inner 1/2-inch flat portion of the electrode.

Also to be noted especially is the fact that all the measurements made on the paper samples for density, air resistance, thickness, and so forth were made in a constant humidity room under standard conditions of temperature and humidity.

The chemical properties of paper are especially important in determining its suitability for use as an insulator. These are affected by the presence of electrolytic impurities, resinates, acid salts, lignin, and ash.

The presence of electrolytic impurities is very objectionable in electrical insulating papers. Among the impurities are chlorides, hydroxides, sulfates, and sulfides which result from cooking and bleaching operations performed on paper pulp during the paper-making process. Another common source of impurities is the excessive reuse of paper machine water as an economic measure.

Rosin-sized papers should not be used and the paper machine should be well cleaned before making a run of insulating paper, or better still have a separate paper machine for such papers.

The ability to withstand temperature extremes without embrittlement is dependent almost entirely upon the absence of acids or acid salts. A very minute amount of an acid salt would affect the high temperature characteristics of the paper, and it might not have an effect upon the conductivity. The acidity, rather than the copper number of the unaged paper, is the significant criterion of heat stability. The increase in copper number is a measure of the amount of physical degradation produced by heat treatment.

The ash content of the paper is an indication of the contamination by clay, fillers, inorganic salts, and so forth. High ash content affects electrical properties adversely. The loss factor of dry unbeaten pulp kraft pulp increases with ash content.

The presence of lignin increases the loss factor, the effect being most noticeable at high frequencies.

The manufacture of capacitor paper requires very prolonged beating so as to eliminate any weak spots as far as possible. Several thin sheets are used in capacitors. Also the dielectric loss must be kept low.

The manufacture of cable paper is quite different. The beating treatment is very slight so that the fibers are not shortened very much, and they retain their original diameter. The fibers should not be gelatinized or fibrillated excessively because this would destroy the porosity. The paper is gently calendered on the machine.

Cable papers are characterized by low density, high strength, and good stretch. Since the papers are impregnated with oil, they should be porous. Dielectric loss and heating effects must be kept to a minimum.

Hydraulic pressure in the press does not break up the fibers like beating does; nevertheless, either an increase in hydraulic pressure in the press or longer beating time, results in increased density of the paper and in increased air-resistance, and the dielectric breakdown increases with increase of density and with increase of air-resistance.

The hydraulic pressure applied is very important since it affects the density of the sheet and its strength; to some degree high hydraulic applied pressure is equivalent to longer beating. It is doubtless true that the over-all effect of beating is that it inhibits a better bonding of the fibers. Air resistance is generally recognized as a measure of the degree of bonding of the fibers.

Table IV

Density, Grams Per Cubic Centimeter	Air Resistance, Seconds	Breakdown Volts Per Mil 60 Cycle (Effective)
0.772	49.80	1,717
0.529	4.82	1,210

Figure 5 is, of course, the crux of the whole matter. The authors feel that the great care exercised at every step is justified fully by the curve obtained. No very marked gain in dielectric strength is obtained by prolonged beating or by higher hydraulic pressures. While the gain might be important in some cases, it would not be significant in others. Also, a sharp drop is observed in dielectric strength below 200 seconds air resistance.

At some future time it is hoped that a series of tests may be made at constant air resistance and varying density.

It should be mentioned, however, that in the original thesis[6] from which the present paper is prepared, an attempt was made to secure constant air resistance and varying density. Two sets of papers were prepared and the data obtained are given in Table IV.

References

1. The True Breakdown Strength of Electric-Cable Insulating Papers, **Paul Cloke, Burt Bates.** *AIEE Transactions,* volume 67, part II, 1948, pages 1191–95.

2. The Dielectric Breakdown Strength of Impregnated Paper Insulation, **Clifton Whitten.** *Master's Thesis,* University of Maine, Orono, Maine, 1950.

3. Laboratory Processing of Pulp (Beater Method). *Standard Number T200,* The Technical Association of the Pulp and Paper Industry, New York, N. Y., 1945.

4. Forming Handsheets for Physical Tests of Pulp. *Standard Number T205,* The Technical Association of the Pulp and Paper Industry, New York, N. Y., 1950.

5. Physical Testing of Pulp Handsheets. *Standard Number T200,* The Technical Association of the Pulp and Paper Industry, New York, N. Y., 1946.

6. A Study of the Dielectric Properties of Oil Impregnated Paper as Affected by Density and Permeability, **K. K. Khandelwal.** *Master's Thesis,* University of Maine, Orono, Maine, 1951.

Discussion

William A. Del Mar (Phelps Dodge Copper Products Corporation, Yonkers, N.Y.): The dielectric strength of oil-impregnated paper and of cables insulated with that material, as affected by the air resistance and specific gravity, is not understood completely as many contradictory data have seemed to defy analysis. It is only by taking one variable at a time that complete understanding will be achieved, and this is what the authors have tried, using paper of variable air resistance and constant specific gravity.

A cable, however, is not merely insulated with impregnated paper but a composite of impregnated paper and oil channels and, under certain circumstances, the oil channels may be the determining factor in the breakdown voltage. This was shown clearly in a series of papers by J. B. Whitehead[1-3] which describe tests on conductors $2^{1}/_{2}$ inches and 1.0 inches in diameter with 60 and 80 mils of paper respectively giving diameter ratios of 1.05 and 1.15 respectively. In these tests, the cables with paper of high specific gravity and high Gurley air resistance gave lower breakdown voltage than cables made with paper of low specific gravity and low Gurley air resistance. Dr. Whitehead explained this by pointing out that the paper of high specific gravity has a higher dielectric constant and thereby throws a greater stress on the oil in the channels.

Does this explain why Cloke and Khandelwal say, "Cable papers are characterized by low density," in spite of the fact that the burden of their paper is to show the greater dielectric strength of high-density paper?

Figure 5 indicates rising dielectric strength with increasing air resistance and Figure 3 indicates rising, followed by constant dielectric strength with increasing specific gravity. However, if the specific gravity data of Table II, from which Figure 5 is derived, are plotted against dielectric strength, a scattering of points is obtained which shows no connection between the two quantities. Which are more reliable, the five 60-cycle points of Figure 3 or four points of Table II?

In view of this apparent contradiction in the data, it does not appear that we yet have the much desired separation of air resistance and specific gravity as factors in the dielectric strength of impregnated paper. Tests with variable specific gravities and constant air resistance are now required to complete our understanding.

So far I have distinguished between specific gravity and air resistance, but in what follows the word density will be used to cover both, as in the papers commonly used, these two properties go together.

Returning to Dr. Whitehead's data and particularly to his 1951 paper,[3] the present writer in his discussion[4] pointed out that the density of paper has two opposing effects on the breakdown voltage of a cable and that the net effect of the density may be to favor either high or low breakdown voltage, depending on the following circumstances.

The dielectric strength of impregnated paper alone, as tested by the authors, is due principally to the baffle action of the matted paper fibers against ion discharges. It, therefore, rises with rising paper density.

The dielectric strength of impregnated paper plus oil spaces depends on the ability of the paper to protect the oil spaces from ionization and is, therefore, greater the lower the dielectric constant of the paper. This means that the dielectric strength of the composite insulation falls with rising paper density. Dr. Whitehead's tests used cables with diameter ratios of 1.05 and 1.15, and, therefore, the stress was substantially uniform throughout and at a level at which the oil ionized.

If a greater insulation thickness had been used so as to obtain more practical diameter ratios, say between 2 and 3, and if the same stress had been applied to the inner layers, I expect that the resistance to outward penetration of ionic streamers from the inner layers would have been greater the higher the density of the un-ionized outer layers. Here the higher dielectric strength of the outer paper resulting from its higher density, would counterbalance the oil ionization effect.

In high-voltage cables the diameter ratio is usually between 2 and 3 and the dielectric strength of the paper itself, therefore, predominates. This is the important case where such data as given by the authors can be of value to point the way to obtaining cables of greater breakdown voltage. The impulse curve of Figure 3 is particularly interesting as it explains the well-known fact that high-density paper must be used to obtain high-impulse strength. In this case, oil ionization has not time to form and breakdown is entirely a function of the baffle effect. Hence the decided upward swing of the curve.

REFERENCES

1. THE DIELECTRIC STRENGTH AND LIFE OF IMPREGNATED-PAPER INSULATION—I. THE INFLUENCE OF THE DENSITY OF THE PAPER, J. B. Whitehead. *AIEE Transactions*, volume 59, 1940, pages 715–20.

2. OIL STRESSES IN IMPREGNATED-PAPER INSULATION, J. B. Whitehead. *Electrical Engineering* (*AIEE Transactions*), volume 64, August 1945, pages 555–57.

3. RADIAL AND TANGENTIAL STRESSES IN IMPREGNATED PAPER INSULATION, J. B. Whitehead. *AIEE Transactions*, volume 70, part I, 1951, pages 56–59.

4. *Discussion* by William A. Del Mar of RADIAL AND TANGENTIAL STRESSES IN IMPREGNATED PAPER INSULATION, J. B. Whitehead. *AIEE Transactions*, volume 70, part I, 1951, pages 61–62.

C. N. Works (Westinghouse Research Laboratories, East Pittsburgh, Pa.): It has been known for some time that insulating papers of the same density did not necessarily have the same dielectric strength. The success of the authors in holding the density of the paper essentially constant while varying the air resistance over a considerable range is a substantial contribution to a better understanding of the factors that control the dielectric strength of insulating papers.

While it is evidently true for the batch of pulp with which the authors worked that the degree of control of dielectric strength with either density or air resistance is limited, there is evidence that this is not true of insulating paper in general, at least not to the extent indicated in Figures 3 and 5 of the paper. For example, oil-impregnated 1/2-mil Kraft capacitor paper stacked to a total thickness of about 5 mils and tested between 2-inch-diameter sharp-edged electrodes has a dielectric strength of 2,275 volts per mil, 60 cycles rms. No attempt was made to eliminate edge effect. This paper has a density of 0.98 and an air resistance of about 5,000 seconds. The test voltage was raised at the rate of about 2,000 volts per second. Oil-impregnated paper used for insulating coils when tested under the same conditions as the Kraft capacitor paper has a dielectric strength of only about 1,200 volts per mil rms. This 3-mil impregnated paper coil insulation has a density of about 0.7 and an air resistance of about 100 seconds. This example is given to show that by varying the air resistance over a wide range and the density from 0.7 to 0.98, a change in dielectric strength of about 2 to 1 was obtained in spite of the fact that the curves of both Figures 3 and 5 reach nearly constant values. The air resistance of the 3-mil paper of 0.7 density is above the minimum value plotted in Figure 5 and the density is only somewhat lower than that of the samples used to obtain the data shown in this figure. This suggests that it would be interesting to measure the dielectric strength of some Kraft capacitor papers of high air resistance using the electrodes of Figure 1.

Table III contains exactly the same information as Figure 3 and therefore could have been omitted. If the air resistance of these samples were added to this table, it would make the printing of this table worth while. Consulting reference 1 of the paper failed to disclose any information on the air resistance of these samples.

Since the title of the present paper is "The Variation at Constant Density of the Dielectric Breakdown of Paper with Air Resistance," Figure 3 is not relevant to this paper but rather is an amendment to the paper of reference 1. However, since this figure is included, it requires some comment. The large value of impulse ratio for paper with a density of 1.0 or greater is unusual. The 60-cycle value of about 2,000 volts per mil rms is no higher than could be expected with 1/4-inch American Society for Testing Materials electrodes since values as high as 2,500 volts per mil rms have been reported on cable paper with a density of 0.95 and it may be assumed in the absence of any data that the air resistance of this paper is typical of that of commercial cable papers of the same density. The maximum impulse stress in an impregnated paper cable is 2,400 volts per mil crest as given by Halperin and Shanklin[1] and the value of 2,600 volts per mil crest for $1^{1}/_{2}$x40-microsecond wave for pressboard[2] caused me to check reference 1 of the paper for the method of measuring the impulse voltages of Figure 3. The oscillograph calibration was not checked by the direct measurement of the impulse voltage by a sphere gap or other means. This leaves some doubt about the accuracy of these impulse voltages. This same question was raised by at least one of the discussers of reference 1 of the paper. This question could be answered definitely with very little trouble by checking the oscillograph calibration against the sparkover of a suitable sphere gap; see AIEE Standard *Number 4*.[3] Also, oil-impregnated paper coil insulation has an impulse strength of about 2,000 volts per mil crest for thin walls.

However, referring to some unpublished tests on flat sheets of 3-mil impregnated paper, an impulse value of 1,850 volts per mil crest was obtained for 2-inch-diameter electrodes for a paper having a density of about 0.7 and an air resistance less than 100 seconds. This paper was below the range of density plotted in Figure 3 of the paper, but it may be noted that the curve is falling rapidly for lower densities. For 1/2-mil Kraft capacitor paper, density 0.98 and air resistance about 5,000 seconds, an impulse value of 5,500 volts per mil crest was obtained on a thickness of about 5 mils (ten sheets). Therefore, the values of impulse strength shown in Figure 3 may well be accurate. This indicates that the impulse strength of the completed cable should not be compared to that of a single sheet of paper. The important comment on Figure 3 is that the authors, after going to so much trouble to separate the effect of air resistance from density, should not have fallen back to graphing dielectric strength as a function of density without regard or even mention of air resistance.

The authors state that the dielectric strength increases with increase of either density or air resistance. Because the authors have separated air resistance from density, it is to be implied that increasing density means increasing density at constant air resistance. The air resistance of the samples listed in Table III of the paper may well vary more than the density varies. One then may argue that the variation of the dielectric strength in Figure 3 is due more to variation in air resistance than to variation in density. In other words, there is no definite evidence in this paper that the dielectric strength will increase with increasing density at constant air resistance. The authors may have data to support their statement, but this is not disclosed in the paper.

In order to clarify this matter, it is hoped that the result of a second series of tests to be made at constant air resistance and varying density as mentioned in this paper will be published in the near future. It is interesting to note that if there is not a misprint in Table IV, when an attempt was made to secure constant air resistance and varying density a 30-per-cent change in density was accompanied by a tenfold change in air resistance. This makes it appear difficult indeed to vary the density of the paper over much of a range and hold the air resistance constant. However, if direct separation of density and air resistance proves too difficult, it should be possible to obtain an indirect separation by selecting densities above and below 0.76 and obtaining air resistance within the range covered in Figure 5.

It would have added to the interest of the paper if the pressure in the press and the beating time for each set were also given in Table II.

It should not be implied that the breakdown data on multiple sheets of impregnated paper given in this discussion is to be compared directly with that for single sheets tested between special electrodes which is the basis for the data given in the paper. To compare the results of dielectric strength tests, all conditions of test, especially the electrode configuration, should be the same. These data were cited for two purposes. First, it is shown that a greater variation in 60-cycle dielectric strength with increasing density and air resistance than appears possible from Figures 3 and 5 may be obtained. Second, since the 60-cycle and impulse values of dielectric strength given in this discussion bracket those of the paper, the values given in the paper appear to be reasonable.

REFERENCES

1. IMPULSE STRENGTH OF INSULATED-POWER-CABLE CIRCUITS, Herman Halperin, G. B. Shanklin. *AIEE Transactions*, volume 63, 1944, pages 1190-97.

2. IMPULSE DIELECTRIC STRENGTH CHARACTERISTICS OF LIQUID-IMPREGNATED PRESSBOARD, T. W. Dakin, C. N. Works. *AIEE Transactions*, volume 71, part I, November 1952, pages 321-25.

3. MEASUREMENT OF TEST VOLTAGE IN DIELECTRIC TESTS. *AIEE Standard Number 4*, November 27, 1942.

Paul Cloke: The aim in this particular paper was to investigate the effect on dielectric breakdown of paper of one variable and, at the request of the reviewers, to give some discussion of the various factors involved. In other words, the approach was from the standpoint of a physical problem without reference to any specific practical application. However, the authors are aware of the excellent contributions of Dr. Whitehead and his associates.

Mr. Works' comments are much appreciated by the authors. In general, I wish to say that the amount of time which might be devoted to such a subject as a graduate thesis is distinctly limited. If one had at his command the vast possibilities of a large research laboratory with adequate personnel, an exhaustive investigation of the many factors involved might have well been attempted and consummated. The authors first wish to state that the statement "the degree of control of dielectric strength with either density or air resistance is limited" was included since it was the comment of a very distinguished engineer and research worker in this field who has had at least 45 years of experience with insulating papers.

Mr. Works' comments on a test with sharp edged electrodes. It is perfectly evident from the work of the authors that it would be unreliable to compare results with sharp edged electrodes with those that were used in this investigation.

Table III of the paper could not be omitted in the opinions, at least, of the reviewers, particularly since it contains information on the deviation. The results of Figure 3, as explained in the paper, refer to previous investigations in which no attempt was made to measure air resistance. However, samples of these papers are still available and could be sent to a research laboratory which is available to know the desired measurements. It will be of interest to Mr. Works, I am sure, to know that the oscillograph calibration was checked in several ways including standard sphere gaps. The authors have made it clear in the paper that it would be desirable to make measurements at constant air resistance with variable density as explained here. This was impossible in the present investigation or in previous ones.

The data in Table IV were included simply for what they were worth. Also the authors are perfectly aware that conditions change when multiple sheets are used.

Design of Unequal-Q Double-Tuned Transformers

SID DEUTSCH
NONMEMBER AIEE

THERE are many instances, in the application of double-tuned transformer coupling, where the primary and secondary Q factors are unequal. When coupling from a crystal converter to the grid of the first intermediate-frequency amplifier, for example, the primary Q is likely to be much lower than the secondary Q. A similar situation exists, if the primary is tuned, when coupling from a transmission line to the grid of a radio-frequency amplifier. An intermediate-frequency interstage transformer, with a damping resistor only across the secondary, may have a high primary Q and low secondary Q.

The unequal-Q case is discussed in the handbooks[1,2] and in various papers.[3,4] However, it is felt that the coverage does not fully present design data and criteria. This paper will present equations and design curves which should reduce the labor required in determining such quantities as the coefficient of coupling, the shape of the response curve, the gain of the stage, the bandwidth, and the phase-shift characteristics of the transformer. Only the small percentage bandwidth, relatively high Q case is discussed.

In the first part of this paper the constant current pentode driven transformer shown in Figure 1(A) is considered. Q is defined, as usual, by X_L/r. In addition, the following relationships are useful when converting to the parallel damping resistor circuit of Figure 1(B)

$$r=\frac{X_0}{Q}=\frac{X_0^2}{R}=\frac{R}{Q^2} \qquad (1)$$

where X_0 is the inductive or capacitive reactance at resonance.

Although the primary and secondary of Figures 1(A) and (B) are coupled by means of mutual inductance, the derived equations are, of course, applicable to other forms of coupling such as the T network of Figure 1(C).

1. Gain versus Frequency

The gain of the stage, Figure 1(A), may be found conveniently by using the re-

---

Paper 52-212, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted December 26, 1951; made available for printing May 19, 1952.

SID DEUTSCH is with Polytechnic Research and Development Company, Inc., Brooklyn, N. Y.

The author is grateful to Franklin Blecher for several useful suggestions concerning the paper.

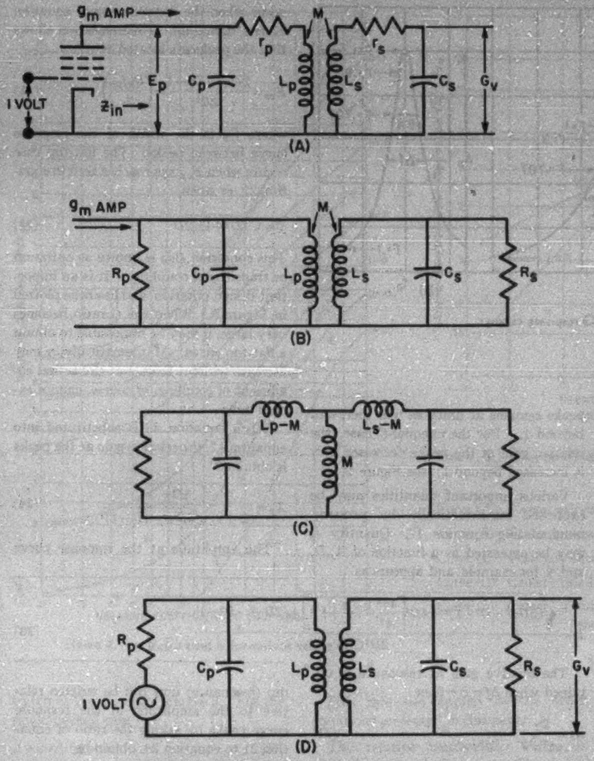


Figure 1. Double-tuned transformer circuits

(A). Constant current pentode driven transformer
(B). Parallel damping resistor arrangement
(C). Equivalent T network
(D). Constant voltage driven transformer

flected impedance technique. However, it is desirable to express all quantities, wherever possible, in terms of R_p, R_s, and Q_p.

Accordingly

$$X_{Lp}-X_{cp}=\frac{R_p}{Q_p}\left(\frac{f}{f_0}-\frac{f_0}{f}\right) \tag{2}$$

where

f is the frequency variable
f_0 is the resonant frequency

A new frequency variable, E, is introduced as follows

$$\sqrt{E}=Q_p\left(\frac{f}{f_0}-\frac{f_0}{f}\right)=\frac{Q_p\Delta f}{f_0} \tag{3}$$

where Δf is the width of the response curve. Equation 3 is valid in the small percentage bandwidth case, and $\sqrt{E}$ becomes the relative curve width. Then

$$X_{Lp}-X_{cp}=(R_p\sqrt{E})/Q_p^2 \tag{4}$$

and

$$X_{Ls}-X_{cs}=(R_s\sqrt{E})/DQ_p^2 \tag{5}$$

where D is the Q ratio

$$D=Q_s/Q_p \tag{6}$$

Also, from equation 1

$$r_p=R_p/Q_p^2 \text{ and } r_s=R_s/(D^2Q_p^2) \tag{7}$$

Critical coupling is defined by

$$K_c=1/\sqrt{Q_pQ_s} \tag{8}$$

If the ratio of actual coefficient of coupling to that of critical coupling is denoted by γ, then

$$X_M=K\sqrt{X_{Lp}X_{Ls}}=(\gamma\sqrt{R_pR_s})/DQ_p^2 \tag{9}$$

In deriving equation 9, it has been assumed that $f\cong f_0$, a step which is valid in the small percentage bandwidth case.

If the foregoing substitutions are made, we emerge with

$$G=(g_mA\sqrt{R_pR_s})/2 \tag{10}$$

where

$$A=\frac{-j2\gamma}{\gamma^2+(1+j\sqrt{E})(1+jD\sqrt{E})} \tag{11}$$

Quantity A is the relative gain variable.

2. Input Impedance and Phase Shift

If the previous approximations are employed, and it is assumed that certain reactive terms are negligibly small at resonance, the input impedance of the transformer appears as

$$Z_{in}=R_p\left[\frac{1+jD\sqrt{E}}{\gamma^2+(1+j\sqrt{E})(1+jD\sqrt{E})}\right] \tag{12}$$

At resonance, Z_{in} becomes

$$Z_{in}=R_p/(1+\gamma^2) \tag{13}$$

and is resistive. The input impedance is also resistive when

$$E=(D\gamma^2-1)/D^2 \tag{14}$$

When $D=1$, it turns out that the input impedance is resistive at the peaks of the response curve as well as at resonance.

With the aid of equation 13, the voltage "rise" across the transformer, at resonance, is given by

$$\frac{G_0}{E_{p0}}=\gamma\sqrt{\frac{R_s}{R_p}} \tag{15}$$

Equation 11 shows that a 90-degree phase shift always exists at resonance. The over-all phase shift at any other frequency, with respect to the phase shift at resonance, is given by

$$\psi_r=\text{arc tan}\,\frac{\sqrt{E}(1+D)}{1+\gamma^2-DE} \tag{16}$$

3. Basic Equations

The basic equations of double-tuned transformer coupling are obtained now by taking the absolute value of equation 11 (absolute value signs are omitted)

$$A=\frac{2\gamma}{\sqrt{(1+\gamma^2-DE)^2+E(D+1)^2}} \tag{17}$$

If this derivation is carried out while interchanging Q_p and Q_s in equations 3 and 6, equations 10 and 17 are obtained

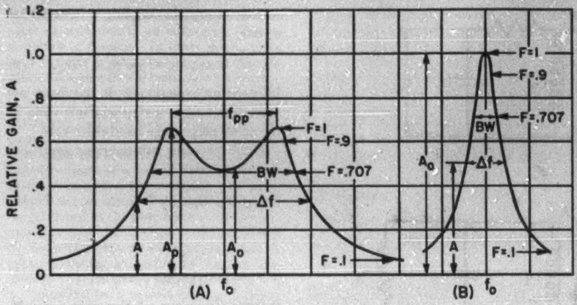


Figure 2. Unequal-Q response curves

(A). When D=8 or 0.125 and $\gamma=4$
(B). When D=8 or 0.125 and $\gamma=1$

again. It is therefore possible to generalize E and D by expressing them in terms of Q_a and Q_b, as follows

$$\sqrt{E}=(Q_a\Delta f)/f_0 \tag{18}$$

and

$$D=Q_b/Q_a \tag{19}$$

Unless otherwise stated, it will be assumed that Q_a is known beforehand, and may be that of the primary or secondary. Q_b is that of the other winding, and may or may not be known beforehand.

The unequal-Q response curve has been plotted, in Figure 2, for $D=8$ or 0.125. The double-peaked curve is for $\gamma=4$, while the single-peaked curve is for $\gamma=1$ (critical coupling). When primary and secondary Q's are equal, a flat-top curve is secured for $\gamma=1$. For $D=8$ or 0.125, however, γ must be increased to 2.015 before a flat-top curve is obtained. Also, for the equal-Q case, the relative gain at the peaks remains at unity as γ is increased beyond 1. For the unequal-Q case, the relative gain at the peaks decreases as γ is increased beyond 1; see Figure 2.

Various important quantities may be expressed mathematically by properly manipulating equation 17. Quantity E may be expressed as a function of A, D, and γ, for example, and appears as

$$E=\frac{\sqrt{(2D\gamma^2-D^2-1)^2+4D^2\left[\frac{4\gamma^2}{A^2}-(\gamma^2+1)^2\right]}+(2D\gamma^2-D^2-1)}{2D^2} \tag{20}$$

The relative gain at resonance is obtained when $\Delta f=E=0$, or

$$A_0=\frac{2\gamma}{\gamma^2+1} \tag{21}$$

The relative gain at resonance is maximum for $\gamma=1$, or critical coupling, and is independent of the Q ratio.

For the double-peaked curve, the peaks occur when the denominator of equation 17 is a minimum. Differentiation shows that the peaks are located at

$$E_{pp}=\frac{2D\gamma^2-D^2-1}{2D^2}=\left(\frac{Q_af_{pp}}{f_0}\right)^2 \tag{22}$$

where f_{pp} is the width of the response curve between peaks. The flat-top case occurs when E_{pp} approaches zero in equation 22, or when

$$\gamma=\sqrt{(D^2+1)/2D} \tag{23}$$

This condition also is known as optimum or transitional coupling. It is an important design criterion, and has been plotted in Figure 3. When the Q ratio becomes very large, it may be impossible to obtain a flat-top curve. Mathematically, γ may increase without limit, but the actual coefficient of coupling, of course, cannot exceed unity.

When equation 22 is substituted into equation 17, the relative gain at the peaks is obtained

$$A_p=\frac{4D\gamma}{(D+1)\sqrt{4D\gamma^2-(D-1)^2}} \tag{24}$$

The amplitude at the response curve dip (resonance) now can be written relative to the amplitude at the response curve peaks by taking the ratio of equation 21 to equation 24, obtaining

$$A_r=\frac{A_0}{A_p}=\frac{G_0}{G_p}=\frac{(D+1)\sqrt{4D\gamma^2-(D-1)^2}}{2D(\gamma^2+1)} \tag{25}$$

If we solve for the relative coefficient of coupling in equation 25, γ appears as

$$\gamma=\frac{1}{A_r}\sqrt{\frac{(D+1)^2(1+\sqrt{1-A_r^2})-2DA_r^2}{2D}} \tag{26}$$

A_r becomes unity when γ yields a flat-top response curve.

4. Bandwidth of Single-Peaked Curve

Although A is the relative gain variable, the relative gain does not reach unity at resonance, for the single-peaked case, unless $\gamma=1$. In general, as equation 21 shows, A_0 will be less than 1. It becomes convenient to express the relative gain in terms of a factor F such that $F=1$ at the peak of the curve. This is illus-

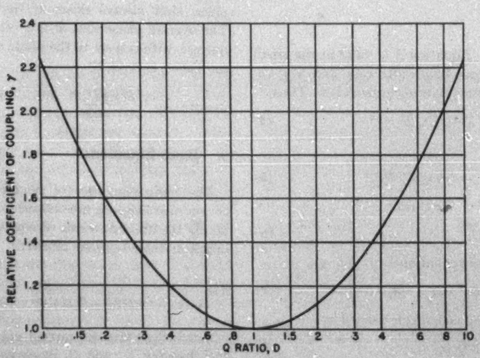


Figure 3. γ versus D for a flat-top response curve

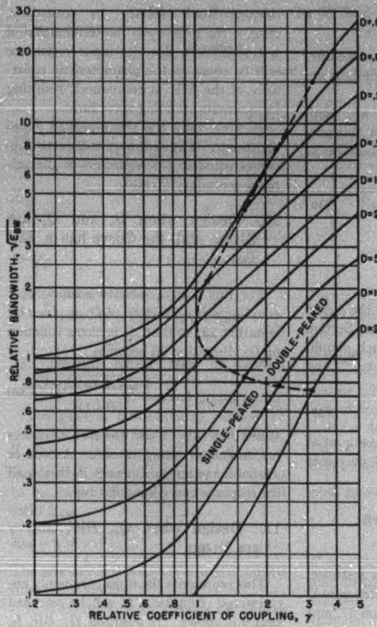


Figure 4. $\sqrt{E_{BW}}$ versus γ for various values of D

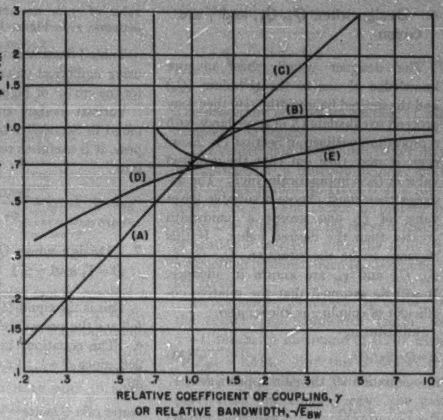


Figure 5. Gain-bandwidth factor versus γ or $\sqrt{E_{BW}}$ under various conditions

Curve (A): Center gain-bandwidth factor versus γ when $D=1$ and $\gamma \leq 1$
Curve (B): Center gain-bandwidth factor versus γ when $D=1$ and $\gamma > 1$
Curve (C): Peak gain-bandwidth factor versus γ when $D=1$
Curve (D): Center gain-bandwidth factor versus $\sqrt{E_{BW}}$ when $\gamma=1$
Curve (E): Center gain-bandwidth factor versus $\sqrt{E_{BW}}$ when the response curve has a flat top

trated in the curves of Figure 2. For the single-peaked case

$$A=2\gamma F/(\gamma^2+1) \tag{27}$$

If F is set equal to 0.707 and equation 27 substituted into equation 20, the relative bandwidth squared, for the single-peaked curve, appears as

$$E_{BW}=\frac{\sqrt{(D^2+1-2D\gamma^2)^2+4D^2(\gamma^2+1)^2}-(D^2+1-2D\gamma^2)}{2D^2} \tag{28}$$

where

$$\sqrt{E_{BW}}=(Q_aBW)/f_0 \tag{29}$$

If equation 28 is solved for D, we obtain

$$D=\frac{\gamma^2\pm\sqrt{\dfrac{(\gamma^2+1)^2}{E_{BW}}+2\gamma^2(\gamma^2+1)-E_{BW}}}{E_{BW}+1} \tag{30}$$

The plus or minus sign is necessary because, under certain conditions, two values of D may satisfy a given set of γ-E_{BW} values. In general, when imaginary or negative answers are obtained, they must be discarded as being physically unrealizable.

This rule also applies to the other equations presented in this paper.

The relative bandwidth, $\sqrt{E_{BW}}$, is plotted versus γ in Figure 4. The region to the left of the dotted line applies to the single-peaked response curve.

5. Bandwith and Steepness Ratio of Double-Peaked Curve

For the double-peaked case, the relative gain does not reach unity at the peaks unless primary and secondary Q's are equal. In general, A_p will be less than 1. To obtain the relative bandwidth, relative gain is expressed in terms of F, so that $F=1$ at the peaks, and F is set equal to 0.707. Equation 20 then yields

$$E_{BW}=\frac{\sqrt{(2D\gamma^2-D^2-1)^2+2(D+1)^2[4D\gamma^2-(D-1)^2]-4D^2(\gamma^2+1)^2}+(2D\gamma^2-D^2-1)}{2D^2} \tag{31}$$

The region to the right of the dotted line in Figure 4 is a plot of $\sqrt{E_{BW}}$ versus γ for the double-peaked case.

As an aid in visualizing the shape of the response curve, the steepness ratio S is introduced as the ratio of curve width at $F=0.1$ to the curve width at $F=0.9$. For the ideal case, that of a rectangular response curve, S is equal to unity. For a single-tuned response curve, S is equal to 20.5. If γ is written in terms of A_r through the use of equation 26, the steepness ratio of a double-peaked response curve becomes

$$S=\sqrt{\frac{\sqrt{1+\dfrac{100A_r^2-1}{1-A_r^2}}+1}{\sqrt{1+\dfrac{1.235A_r^2-1}{1-A_r^2}}+1}} \tag{32}$$

From the fact that S is independent of D in equation 32, one may infer that the shape of a double-peaked response curve is completely specified once A_r is given.

The value of S for the flat-top curve ($A_r=1$) is 4.53. The response curve sides become steeper as A_r is reduced.

6. Design when Q_a, Q_b, and γ are Given

The designer usually has advance knowledge of one of the Q values (Q_a) and the desired bandwidth. He then may proceed to calculate Q_b in accordance with a procedure outlined in Section 11. However, it may turn out that the required value of Q_b is impractically high. The designer then must choose a suitable lower value of Q_b and accept a bandwidth greater than the desired value. If this happens to be the situation, Q_a, Q_b, f_0, C_p, C_s, and g_m are known in advance. It will be assumed that the relative coefficient of coupling is known also.

Substituting

$$R = Q/(2\pi f_0 C) \tag{33}$$

into equation 10, the gain appears as

$$G = \frac{g_m A}{4\pi f_0}\sqrt{\frac{Q_a Q_b}{C_p C_s}} \tag{34}$$

The following procedure may be employed now to find D, BW, f_{pp}, G_0, G_p, and K:

D: The Q ratio is found immediately from equation 19.

BW: One must first determine, from equation 23 or Figure 3, whether the response curve has one or two peaks. Since D and γ are known, the value of E_{BW} may be found from equation 28 (single-peak), or equation 31 (double-peak), or from Figure 4. Equation 29 then will yield the bandwidth.

f_{pp}: If two peaks exist, equation 22 may be solved for f_{pp}.

G_0: Equation 21 will supply the value of A_0 to be substituted into equation 34.

G_p: The gain at the peaks follows from equation 25.

K: The necessary coefficient of coupling is given by

$$K = \gamma/\sqrt{Q_a Q_b} \tag{35}$$

If it is found that K is impractically large, it will be necessary to assume a smaller value of γ or use an equivalent T or π network, as in Figure 1(C).

Table I illustrates the design procedure, using numerical values. The designs are for the curves of Figure 2.

Various design conditions are compared in Sections 7 to 10. For this purpose, it is useful to rearrange equation 34 into

$$\text{“gain-}Q\text{” factor} = \frac{2\pi f_0 G}{g_m}\sqrt{\frac{C_p C_s}{Q_a Q_b}} = \frac{A}{2} \tag{36}$$

7. Design when Q_a is Given, $D=1$, and $\gamma \leqq 1$

This is the equal-Q case, and the design follows the procedure outlined in Section 6. The equations become much simplified, since $D=1$. From equations 21 and 36

$$\text{center gain-}Q\text{ factor} = \gamma/(\gamma^2+1) \tag{37}$$

The available center gain is maximum at critical coupling ($\gamma=1$), and decreases rapidly for smaller values of γ.

8. Design when Q_a is Given, $D=1$, and $\gamma > 1$

The response curve, in this instance, will have two peaks. The center gain-Q factor again is given by equation 37, and shows that the available center gain rapidly decreases as γ is made larger than 1.

For the equal-Q case, equation 24 shows that A_p remains constant at unity. From equation 36, therefore, the peak gain-Q factor is equal to 0.5 for $1 < \gamma < \infty$.

9. Design when Q_a and Q_b are Given and $\gamma = 1$

This case is important because it represents maximum transfer of energy, at resonance, from primary to secondary. Also, if the primary is the termination of a transmission line, its input resistance will be equal to the characteristic resistance of the line (at resonance), resulting in minimum standing wave ratio.

If $\gamma=1$, only the single-peaked curve can occur. The center gain-Q factor becomes equal to 0.5 regardless of the primary-secondary Q ratio.

10. Design when Q_a and Q_b are Given and the Curve has a Flat Top

For this design, relative coefficient of coupling γ is a function of D, as given by equation 23, so that A_0 is also a function of D. Equation 36 becomes

$$\text{center gain-}Q\text{ factor} = \frac{\sqrt{2D(D^2+1)}}{(D+1)^2} \tag{38}$$

When $D=1$, $\gamma=1$ and the factor reaches its maximum value of 0.5. A loss in available center gain occurs if $D \neq 1$ and the response curve has a flat top.

11. Design when Q_a, BW, and γ are Given

This represents the most commonly encountered situation. It will be assumed that γ is known. In addition, Q_a, BW, f_0, C_p, C_s, and g_m are known in advance.

Substituting equations 19 and 29 into equation 34, the gain appears as

$$G = \frac{g_m A}{4\pi BW}\sqrt{\frac{DE_{BW}}{C_p C_s}} \tag{39}$$

The following procedure may be employed to find D, Q_b, f_{pp}, G_0, G_p, and K:

D: The relative bandwidth is given by equation 29. When γ and $\sqrt{E_{BW}}$ are known, it is possible to employ Figure 4 to determine whether the response curve will have one or two peaks. Figure 4 also will warn the designer when his set of γ-E_{BW} values is physically unrealizable. In the latter event, it will be necessary to assume a different set of given conditions.

If the curve is single-peaked, equation 30 or Figure 4 will yield the value of D. If the curve is double-peaked, equation 31 may be solved for D, but a fourth-degree equation results. For ordinary accuracies, Figure 4 should suffice.

Q_b: The value of Q_b, of course, now may be found from Equation 19.

f_{pp}, G_0, G_p, and K: These quantities may be determined as in Section 6.

The design procedure is illustrated in Table II, using numerical values. The designs are for the curves of Figure 2.

To compare various design conditions in Sections 12 to 15, equation 39 is rearranged into

Table I. Design when Q_a, Q_b, and γ are Given

Quantity	Example 1 Numerical Value	Example 1 How Obtained	Example 2 Numerical Value	Example 2 How Obtained
γ	4	given	1	given
Q_a	5	given	5	given
Q_b	40	given	40	given
f_0	30 megacycles	given	30 megacycles	given
C_p	10 micromicrofarads	given	10 micromicrofarads	given
C_s	10 micromicrofarads	given	10 micromicrofarads	given
g_m	5,000 micromhos	given	5,000 micromhos	given
D	8	equation 19	8	equation 19
Number of peaks	two	equation 23 or Figure 3	one	equation 23 or Figure 3
E_{BW}	3	equation 31 or Figure 4	0.0745	equation 28 or Figure 4
BW	10.39 megacycles	equation 29	1.638 megacycles	equation 29
E_{pp}	1.492	equation 22		
f_{pp}	7.33 megacycles	equation 22		
A_0	0.47	equation 21	1	equation 21
G_0	8.8	equation 34	18.72	equation 34
A_p	0.711	equation 25		
G_p	12.38	equation 25		
K	0.2828	equation 35	0.0707	equation 35

Table II. Design when Q_a, BW, and γ are Given

Quantity	Example 1		Example 2	
	Numerical Value	How Obtained	Numerical Value	How Obtained
γ	4	given	1	given
Q_a	5	given	5	given
BW	10.39 megacycles	given	1.638 megacycles	given
f_0	30 megacycles	given	30 megacycles	given
C_p	10 micromicrofarads	given	10 micromicrofarads	given
C_s	10 micromicrofarads	given	10 micromicrofarads	given
g_m	5,000 micromhos	given	5,000 micromhos	given
E_{BW}	3	equation 29	0.0745	equation 29
Number of peaks	two	Figure 4	one	Figure 4
D	8	Figure 4	8	equation 30 or Figure 4
Q_b	40	equation 19	40	equation 19

E_{pp}, f_{pp}, A_0, G_0, A_r, G_p, and K are found as in Table I.

"gain-bandwidth" factor

$$= \frac{2\pi G\, BW\sqrt{C_pC_s}}{g_m} = \frac{A\sqrt{DE_{BW}}}{2} \qquad (40)$$

12. Design when BW is Given, $D=1$, and $\gamma \leq 1$

The general procedure of Section 11 must be modified slightly now, since Q_a is not known but only the fact that $Q_a = Q_b$ is given. Since $D=1$ and γ is assumed to be known, however, equation 28 will yield E_{BW}

$$E_{BW} = \sqrt{2(\gamma^4+1)} - (1-\gamma^2) \qquad (41)$$

The primary-secondary Q now is obtained from equation 29.

It is instructive to express the center gain-bandwidth factor as a function of γ. When equations 21 and 41 are substituted into equation 40, we obtain

center gain-bandwidth factor

$$= \frac{\gamma\sqrt{\sqrt{2(\gamma^4+1)} - (1-\gamma^2)}}{\gamma^2+1} \qquad (42)$$

The factor is shown, in curve A of Figure 5, as a function of γ. When $\gamma=1$, the factor reaches its maximum value of 0.707. As γ approaches zero, the center gain-bandwidth factor rapidly approaches zero.

13. Design when BW is given, $D=1$, and $\gamma>1$

The curve for this case is of the double-peaked variety. Since $D=1$ and γ is assumed to be known, equation 31 simplifies to

$$E_{BW} = \gamma^2 + 2\gamma - 1 \qquad (43)$$

and the primary-secondary Q, in turn, may be found.

As in Section 12, the center gain-bandwidth factor may be expressed as a function of γ, and appears as

center gain-bandwidth factor

$$= \frac{\gamma\sqrt{\gamma^2+2\gamma-1}}{\gamma^2+1} \qquad (44)$$

This relationship is plotted as curve B of Figure 5. The gain factor reaches a maximum value of 1.12 when γ is approximately equal to 3. This represents considerable improvement over the 0.707 figure obtained when $\gamma=1$, but the response curve has a large peak to either side of center for $\gamma=3$.

If the designer is interested in peak gain rather than center gain, equations 24, 40, and 43 will yield

peak gain-bandwidth factor

$$= \frac{\sqrt{\gamma^2+2\gamma-1}}{2} \qquad (45)$$

Curve C of Figure 5 shows that the peak gain factor rises without limit as γ is increased. As usual, one must remember that the increased gain at the peaks is accompanied by a large center dip, and that the maximum possible value of K is 1.

14. Design when Q_a and BW are Given, and $\gamma=1$

Since the response curve must have a single peak for $\gamma=1$, equation 30 is applicable, and simplifies to

$$D = \frac{1 \pm \sqrt{\frac{4}{E_{BW}} + 4 - E_{BW}}}{E_{BW}+1} \qquad (46)$$

The procedure outlined in Section 11 will complete the design.

The center gain-bandwidth factor may be expressed as a function of E_{BW}, since $A_0=1$ and equation 46 gives D as a function of E_{BW}. The factor is plotted in curve D of Figure 5. This curve is mainly of academic interest, since the design is uniquely determined once Q_a, BW, and γ are given.

Nevertheless, as shown by curve D, the center gain-bandwidth factor, for $\gamma=1$, reaches a maximum of 0.707 when $\sqrt{E_{BW}}=1.414$. But this is simply the equal-Q case, or $D=1$. As the design departs from the equal-Q case, the gain factor decreases.

15. Design when Q_a and BW are Given and the Curve has a Flat Top

For a flat-top response curve, γ is given as a function of D by equation 23. When this expression is substituted in place of γ in equation 28, we emerge with the simple form

$$E_{BW} = \frac{(D+1)^2}{2D^2} \qquad (47)$$

or

$$D = \frac{1}{\sqrt{2E_{BW}} - 1} \qquad (48)$$

When A_0 and D are written in terms of

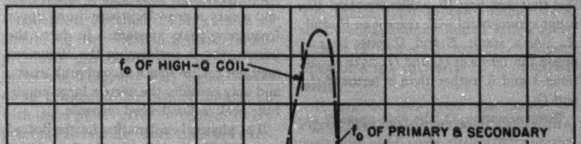


Figure 6. Unequal-Q response curves

—— Correct alignment. — — — Detuned

E_{BW} by means of equations 21, 23, and 48, equation 40 becomes

center gain-bandwidth factor

$$= \sqrt{\frac{1}{E_{BW}}+1-\sqrt{\frac{2}{E_{BW}}}} \qquad (49)$$

The factor is plotted, versus $\sqrt{E_{BW}}$, in curve E of Figure 5. The minimum value of the gain factor occurs at $\sqrt{E_{BW}}=1.414$, or when primary and secondary Q's are equal. The factor approaches a maximum of 1 when D approaches zero or infinity. The reason for the improvement in gain with unequal Q's is that, for given values of Q_a and BW, the flat-top case involves an increase in γ as the primary-secondary Q ratio departs from unity.

16. Constant Voltage Driven Transformer

In Figure 1(D) a double-tuned transformer is driven by a constant voltage source. This situation occurs, for example, when the primary is the termination of a transmission line. If the line is matched at the source, R_p becomes equal to the characteristic resistance of the line.

Using the reflected impedance technique, and applying equations 2 to 9, we find that the gain is given by

$$G=\frac{A}{2}\sqrt{\frac{R_s}{R_p}} \qquad (50)$$

Since A is the relative gain variable, as before, most of the foregoing discussion of this paper applies to the constant voltage driven transformer also. An important point of departure exists, however, in the fact that equation 50, unlike equation 10, is not symmetrical with respect to R_p and R_s. As a result, E and D must be expressed in terms of Q_p and Q_s, as in equations 3 and 6, rather than in terms of Q_a and Q_b.

From equation 33, the gain appears as

$$G=\frac{A}{2}\sqrt{\frac{C_pQ_s}{C_sQ_p}} \qquad (51)$$

or the gain-Q factor becomes

$$\text{gain-}Q\text{ factor}=G\sqrt{\frac{C_sQ_p}{C_pQ_s}}=\frac{A}{2} \qquad (52)$$

For given values of Q_p, Q_s, C_p, C_s, and γ, the discussion of Sections 6 to 10 is applicable.

When equations 19 and 29 are substituted into equation 51, the gain becomes

$$G=\frac{Af_0}{2Q_pBW}\sqrt{\frac{C_pDE_{BW}}{C_s}} \qquad (53)$$

or the gain-bandwidth factor may be written as

gain-bandwidth factor

$$=\frac{G\,BW\,Q_p}{f_0}\sqrt{\frac{C_s}{C_p}}=\frac{A\sqrt{DE_{BW}}}{2} \qquad (54)$$

For given values of Q_p, BW, f_0, C_p, C_s, and γ, sections 11 to 15 are applicable.

17. Alignment of Unequal-Q Transformer

To illustrate the peculiarities of unequal-Q alignment, a typical overcoupled response curve is presented in Figure 6.

The solid-line curve represents correct alignment, with primary and secondary tuned to approximately the same resonant frequency. The dotted-line curve will be obtained when the low-Q coil resonates, by a specific amount, at a lower frequency than that of the high-Q coil. (Of course, a mirror image of the dotted curve will occur if the low-Q coil resonates at a higher frequency than that of the high-Q coil.)

The alignment procedure may become very confusing because, when the inductance or capacitance of the lower-Q coil is increased, one would expect the lower frequency peak to reach a greater amplitude than the upper frequency peak. Instead, exactly the reverse occurs. The lower frequency peak almost disappears, and a very narrow large-amplitude upper frequency peak appears. In fact, the person who is performing the alignment may not realize that a second peak exists, and may consider the narrow large-amplitude peak as the desired response.

The physical explanation for the dotted curve of Figure 6 lies in the impedance that is reflected by one coil into the other. At the lower amplitude peak, the low-Q coil reflects resistance into the high-Q coil, which is at a considerable distance from resonance at this point. At the higher amplitude peak, the low-Q coil reflects capacitive reactance into the high-Q coil, producing resonance without appreciably lowering its Q.

The use of a sweep signal generator and oscilloscope for alignment purposes is almost mandatory when the relative coefficient of coupling is greater than 1. When γ is less than 1, only a single peak should occur, so that the primary and secondary may be peaked up by means of a fixed-frequency signal generator. When γ is greater than 1, the alignment should be visually aimed at the creation of a symmetrical response curve.

Conclusions

From the foregoing discussion, one may conclude that:

1. When primary and secondary Q's are known beforehand, the center gain-Q factor reaches a maximum value of 0.5 at critical coupling ($\gamma=1$).
2. When primary and secondary Q's are known beforehand, the peak gain-Q factor will always be less than 0.5 unless primary and secondary Q's are equal. In the latter event, the factor will be equal to 0.5 for $1<\gamma<\infty$.
3. When the Q_a or Q_p value and the bandwidth are known beforehand and the response curve is to have a single peak, the center gain-bandwidth factor is maximum when the transformer is designed for a flat-top response curve.
4. When the Q_a or Q_p value and the bandwidth are known beforehand and the response curve is allowed to have two peaks, the center gain-bandwidth factor will increase (but not indefinitely) as γ is increased. The peak gain-bandwidth factor, however, will increase without limit as γ is increased.
5. When γ is greater than 1, the transformer should be aligned for maximum symmetry with the aid of a sweep signal generator.

References

1. Radio Engineers' Handbook (book), **Frederick E. Terman.** McGraw-Hill Book Company, Inc., New York, N. Y., 1943, pages 154–63.

2. Reference Data for Radio Engineers (book). Federal Telephone and Radio Corporation, New York, N. Y., third edition, 1949, pages 119–24.

3. Exact Design and Analysis of Double- and Triple-Tuned Band-Pass Amplifiers, **Milton Dishal.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 35, number 6, June 1947, pages 606–26.

4. Two-Mesh Tuned Coupled Circuit Filters, **C. B. Aiken.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 25, number 2, February 1937, pages 230–72.

No Discussion

Impulse Dielectric Strength Characteristics of Liquid-Impregnated Pressboard

T. W. DAKIN
ASSOCIATE MEMBER AIEE

C. N. WORKS
MEMBER AIEE

Synopsis: The impulse dielectric strength of pressboard impregnated with transformer mineral oil and transformer Askarel and a variety of other liquids having differing chemical constitution is reported. It is emphasized that the decreasing dielectric strength with increasing multiple cycles of alternating current is responsible for the high impulse ratio of mineral oil–impregnated pressboard and that this effect is diminished greatly in the case of transformer Askarel, which has a low impulse ratio. Investigation of polarity effects with different electrodes has indicated that breakdown of the Askarel-impregnated pressboard is initiated in the liquid at the edge of the positive electrode.

THE impulse dielectric strength of insulating materials is of interest where these materials are to be applied in electric apparatus which must withstand the transient high-voltage surges produced by lightning. The AIEE has arbitrarily defined an impulse wave which is supposed to be representative of the lightning surges which occur on high-voltage transmission lines. This is the familiar $1^{1}/_{2}$x40-microsecond wave which reaches its crest starting from zero in $1^{1}/_{2}$ microseconds and decays exponentially to one-half its crest voltage in 40 microseconds. All of the impulse tests described in this paper were made using this standard impulse wave shape. At the present time, it is common practice to prooftest completed electric apparatus, especially transformers, with this impulse voltage, the magnitude of the voltage depending on the voltage class of the apparatus. In spite of the generality of impulse testing of apparatus, there is not a great deal of basic information in the literature on the dielectric strength of a variety of dielectric materials using impulse voltages. Because of its more general use, the dielectric strength of mineral oil with impulse voltage has been covered more adequately in the literature.[1–7] Numerous papers have presented data on the impulse strength of air, of course, notable among them being those by Bellaschi and Teague[8] and Dowell and Foust[3] who present a variety of data with different test conditions. It is the purpose of this paper to describe a series of impulse tests with other liquids, in particular the Askarel type, both by themselves and impregnated into pressboard.

The dielectric breakdown of liquids and solids shows a time lag which depends on the voltage and varies from less than a microsecond to as much as minutes. This has been demonstrated in the familiar voltage-time curves by Bellaschi and Teague[1] and Montsinger[3] who plot on the same graph the breakdown voltage as a function of time to breakdown with d-c impulse voltages and 60-cycle alternating voltage. It is not perfectly clear, however, that the times involving multiple cycles of alternating current are comparable with duration of a single d-c pulse. From their data, it is clear, however, that the breakdown voltage with a single cycle of 60-cycle alternating current is about the same as with a standard $1^{1}/_{2}$x40-microsecond impulse. With multiple cycles of alternating current, the breakdown voltage decreases in the case of mineral oil and mineral oil–impregnated pressboard to a value which is nearly constant after 1 minute.

The ratio of the crest breakdown voltage with the $1^{1}/_{2}$x40-microsecond standard impulse to the crest 60-cycle long-time (usually 1 minute) breakdown voltage is known as the impulse ratio. This has a value of about 2.5 for oil with unsymmetrical electrodes, but a value of about 3.0 for 0.056-inch-thick oil-impregnated pressboard, according to Bellaschi and Teague.[1] This higher breakdown voltage in the case of mineral oil with short pulses gives, of course, a very practical advantage, since transient overvoltage pulses often occur in electric apparatus due to switching or lightning surges. As will be shown from the tests, this advantage seems to be lacking in the case of pressboard impregnated with some common Askarels where the impulse ratio is unexpectedly only slightly more than unity.

Impulse Testing Techniques

In the impulse tests which are reported here, the following procedure was used. The voltage was raised in steps of about 5 to 10 per cent of the expected breakdown voltage, using five impulses at each step. The impulses were applied at intervals of approximately 1 to 2 minutes. The breakdown voltage was taken as the voltage at which breakdown occurred after the first impulse but before the fifth impulse at the same voltage. Breakdown occurred at the crest of the wave or a few microseconds later. The starting voltage was always taken at a value low enough such that several steps of voltage would be taken at which no failure occurred in five impulses. The electrodes are as indicated in the text and figures. In all cases, the tests were conducted in nonmetallic tanks with grounds located reasonably distant. The test cells used are shown schematically in Figures 1(A) and (B). Pressboard samples were thoroughly vacuum-dried and impregnated, and kept dry until tested.

Short-Time 60-Cycle Tests

Montsinger[3] and Bellaschi and Teague[1] have already shown that the short-time 60-cycle strength of mineral oil–impregnated pressboard approaches the impulse strength with $1^{1}/_{2}$x40 full wave impulses. Short-time 60-cycle tests were performed on Askarel-impregnated pressboard and repeated on mineral oil–impregnated pressboard for comparison. These tests were performed with impregnated 0.032-inch-thick pressboard in a conventional oil test cup having 1-inch diameter square edge electrodes as shown in Figure 1(A). The Askarel was the common transformer, Askarel comprising 60-per-cent chlorinated diphenyl and 40-per-cent trichlorbenzene. The voltage was applied from an already energized high-voltage transformer by means of a high-voltage switch. The cycles applied before failure at a steady voltage were counted with a cycle counter or high-speed recorder. Failure times of a few cycles could be observed readily.

It can be seen from the curves of Figure 2 that the dielectric strength with a few cycles of 60-cycle alternating current is nearly the same or slightly greater than the impulse level for both the mineral

---

Paper 52-228, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted March 26, 1952; made available for printing May 13, 1952.

T. W. DAKIN and C. N. WORKS are both with the Westinghouse Electric Corporation, East Pittsburgh, Pa.

Acknowledgment is gratefully made for the helpful co-operation of Dr. George Sommerman, formerly of Northwestern University, Evanston, Ill., and now with Battelle Memorial Institute, Columbus, Ohio, with part of this work, and for the encouragement and advice of Dr. C. F. Hill and Dr. L. J. Berberich, both of Westinghouse Electric Corporation, East Pittsburgh, Pa.

oil- and the Askarel-impregnated pressboard. But in contrast to the mineral oil-impregnated pressboard, the long-time 60-cycle strength of the Askarel-impregnated pressboard is not much lower than the short-time strength. It is this fact which gives the Askarel pressboard a low impulse ratio. The relatively large decrease in dielectric strength of the mineral oil pressboard with multiple cycles of alternating current is believed to be due to the more intense corona in the mineral oil at the electrode edges. Since the dielectric constant of the mineral oil is relatively lower than the impregnated pressboard, the electric field intensity would be higher in the mineral oil than in the corresponding case of the Askarel where the liquid and the impregnated pressboard have nearly the same dielectric constant. Puncture holes were always located at the electrode edges with multiple cycles of alternating current. This, of course, does not explain the relatively low impulse level of the Askarel-impregnated pressboard, but only explains its low impulse ratio.

The Effect of Liquid Type on the Impulse Dielectric Strength

The impulse and 60-cycle long-time (1 minute) dielectric strength of transformer Askarel is compared with mineral oil in Figure 3 as a function of gap spacing for small gaps with 1-inch diameter square edge electrodes. The impulse ratio for the Askarel is only 1.38 as compared with 2.00 for the mineral oil. The latter value is somewhat lower than has been previously reported by some authors, but this is probably due to the higher 60-cycle value reported here for the dielectric strength of the purified mineral oil tested here. Purification of the liquid increases the 60-cycle dielectric strength but does not seem to affect the impulse strength as much.

The comparatively low impulse strength of the Askarel is more pronounced when it is impregnated into pressboard. In Figure 4, the dielectric strength of oil- and Askarel-impregnated pressboard are compared. Here it can be seen that the Askarel-impregnated pressboard has an appreciably lower dielectric strength with impulse voltages than oil-impregnated pressboard, whereas the reverse is true in the case of 60-cycle voltages. The impulse ratio in this case is only 1.2 for the Askarel but 2.33 for the oil. Impregnation into pressboard has raised the impulse ratio of oil, but reduced the ratio of the Askarel. This rather surprising effect means that the time delay for the Askarel breakdown is very short compared to the time delay for mineral oil breakdown.

This result seemed to be associated with the chemical structure difference between the liquids. Therefore, a series of liquids of differing and similar chemical structure were impulse-tested in pressboard to discover what chemical factor was involved. For this series of tests, a thin pressboard was selected because of the small volume of liquid required in testing it. The results of these tests are summarized in Table I.

These conclusions can be drawn from an examination of Table I. All the halogenated compounds with the exception of the perfluorinated ones and the cold high-viscosity Aroclor 1254 have low impulse strength. The chlorinated aliphatic compounds seem to have exceptionally low impulse strength and are very prone to flash over on the surface. This is even true of the completely chlorinated aliphatic liquids, both of which have low dielectric constants and very low power factors, which would seem to eliminate the possibility that a high power factor contributes to the low impulse strength.

Viscosity here seems to be a factor in the impulse strength inasmuch as the high-viscosity Aroclor 1254 had a high impulse strength, but reduction in the viscosity either by raising the temperature or dilution with mineral oil produced a lower impulse strength.

Small amounts of additives which might be expected to combine with acid or halogen in the transformer Askarel

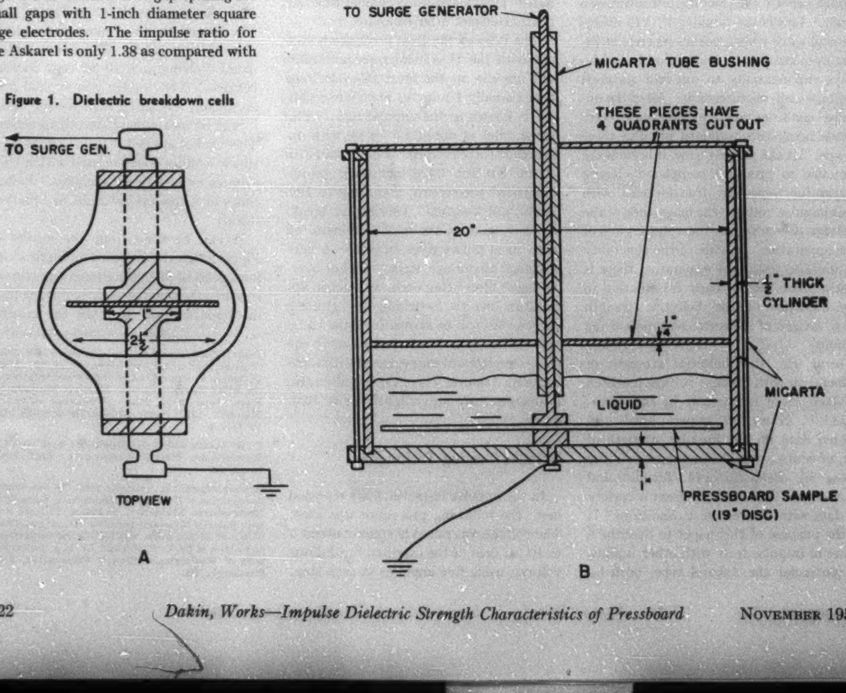


Figure 1. Dielectric breakdown cells

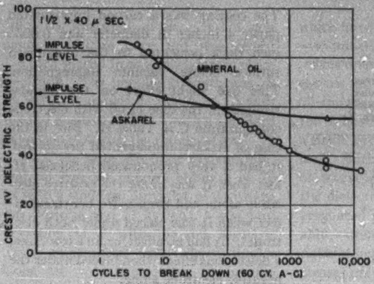


Figure 2. Time to breakdown impregnated pressboard (1/32 inch)

had little or no effect on the impulse strength. Mixtures of the transformer Askarel with high percentage of mineral oil and other hydrocarbons have higher impulse strengths intermediate between that of the Askarel and of the additive.

The perfluoro compounds are notable in that they are the only halogenated compounds which have impulse strengths approaching mineral oil.

Among the nonhalogenated compounds tested, only the phosphate compounds had low impulse strengths. Viscosity in this series does not seem to be a factor, since the low-viscosity ethyl benzene has an impulse strength close to mineral oil. Since these liquids possess quite a range of vapor pressure with nearly equal impulse strength, that is apparently not a factor. Dielectric constant, and the high power factor associated with high dielectric constant, also does not seem to be a large factor as can be seen from the impulse strengths obtained with the series of mixtures of nitro and ethyl diphenyl.

Attempts were made to discover if hydrochloric acid evolved from the Askarel was responsible for the low impulse strength. Pressboard impregnated with mineral oil and tested in mineral oil saturated with this gas was impulse-tested with no appreciable decrease in impulse strength. Addition of small amounts of acetic acid and benzoyl chloride to mineral oil failed to lower its impulse strength in pressboard appreciably. Benzoyl chloride was added with the intention of testing the effect of a compound with a very reactive halogen.

Effect of Pressboard Treatment on the Impulse Strength

Inasmuch as the impulse ratio of the Askarel impregnated into pressboard is lower than with the Askarel liquid alone, while the converse is true with mineral oil, it was thought that modifications of the pressboard might affect the impulse strength.

Numerous experiments were tried to modify the cellulose in the pressboard, such as superdrying, acetylating the fiber surfaces, treating the fiber surfaces with a silicone film, and swelling the cellulose, all of which had slight effect on the impulse strength of the Askarel-impregnated pressboard.

When the pressboard was preimpregnated with a thermosetting resin, however, the impulse strength in the Askarel could be raised considerably. This, in principle, had the effect of excluding the Askarel from the pores of the pressboard. To get the highest impulse strength, the resin monomer had to be vacuum-impregnated into the pressboard and the Askarel subsequently vacuum-impreg-

Table I. Impulse Dielectric Strength of 0.032-Inch Impregnated Pressboard

Impregnating Liquid	Average Impulse Strength, Kv
Mineral Oil	82
Transformer Askarel and Additives	
Transformer Askarel	64
Trichlorbenzene	62
75% transformer Askarel+25% mineral oil	75
50% trichlorbenzene+50% mineral oil	65
50% trichlorbenzene+50% ethyl diphenyl	68
Transformer Askarel+2% polystyrene	67.5
Transformer Askarel+0.5% quinone	62.5
Transformer Askarel+0.5% α naphthyl amine	64
85% transformer Askarel+15% dibutyl phthalate	55
94% transformer Askarel+6% diallyl phthalate	67
95% transformer Askarel+5% vinyl carbazole	65
Capacitor Askarel and Additives	
Capacitor Askarel (chlorinated diphenyl, Aroclor 1254)	
at 25 degrees centigrade	81
at 60 degrees centigrade	67
50% capacitor Askarel+50% mineral oil	72
Miscellaneous Halogenated Compounds	
Fluorinated	
p-Fluoro toluene	69
Benzotrifluoride	55
Fluorocarbon C_7F_{16}	80
Fluorocarbon C_8F_{18}	79
Chlorinated	
Tetrachlorethane	55
Chlorinated benzotrifluoride	65
O-Chloroaniline	65
Hexachlorbutadiene	50 kv (creepage)
Hexachlorbutadiene (60%), mineral oil (40%)	70 65 (creepage)
Hexachlorpropylene	50
Brominated	
Bromobenzene	65
Nonhalogenated Compounds	
Ethyl benzene (boiling point 136.2 degrees centigrade)	79
Ethyl diphenyl	75
75% ethyl diphenyl+25% nitro diphenyl ($\epsilon'=5.18$)	73
50% ethyl diphenyl+50% nitro diphenyl ($\epsilon'=8.09$)	75
25% ethyl diphenyl+75% nitro diphenyl ($\epsilon'=11.58$)	70
Tributyl phosphate	69
75% triethyl phosphate+25% tricresyl phosphate	60
75% tributyl phosphate+25% tricresyl phosphate	65
Dow Corning Silicone Oil 500	75

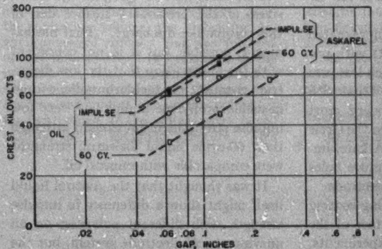


Figure 3 (left). Dielectric strength of Askarel and mineral oil, 1-inch diameter square edge electrodes

Figure 4 (right). Dielectric strength of impregnated pressboard, symmetrical square edge electrodes

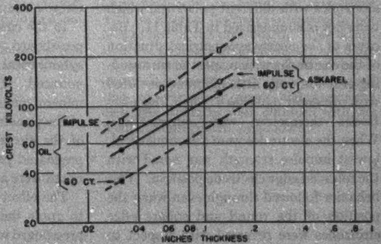
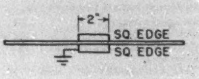
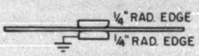
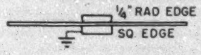
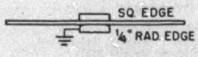
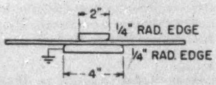
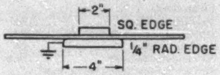


Table II. Impulse Dielectric Strength

1/8-Inch Askarel-Impregnated Pressboard

	Electrode Configuration	Polarity Upper Electrode	Average Impulse Dielectric Strength, Kv
A		+ −	141 140
B		+	163
C		+ −	134 157
D		+ −	155 131
E		+ −	126 152
F		+ −	148 137

nated into the resin-treated pressboard. Impulse strengths about equal to mineral oil–impregnated pressboard could be obtained with a thermosetting hydrocarbon-type resin in the pressboard which swelled in the Askarel.

Polarity Effects with Unsymmetrical Electrodes and the Effect of Stress Concentration at Electrodes

One of the most unusual aspects of the impulse dielectric strength of Askarel-impregnated pressboard is the behavior of this dielectric with unsymmetrical electrodes with different polarities, which behavior is summarized in Table II. Because of the nearly symmetrical situation of the electrodes with respect to surrounding grounds, very little difference is noted with different polarities when the electrodes themselves were symmetrical. When the electrodes were unsymmetrical, lowest impulse strength was noted with the rounded edge electrode positive. This behavior followed through even when the position of the round and square edge electrodes were reversed with respect to ground.

It seems to be indicated by these tests that the breakdown is initiated in the liquid space where the round edge curves away from the surface of the impregnated pressboard. This is consistent with the tests shown in Table II where the lower electrode is larger than the upper. The lowest impulse strength of all configurations was obtained with the top electrode having rounded edges and a positive polarity, Case E. When both electrodes were symmetrical with rounded edges, the stress concentration at the edge would be lower, and, therefore, the dielectric strength was higher.

In the case of mineral oil–impregnated pressboard, the polarity effects with unsymmetrical electrodes are much less pronounced. It has also been noted that with mineral oil–impregnated pressboard, most of the puncture holes are within the circle of the electrode, while with Askarel-impregnated pressboard, all puncture holes noted were at the edge of the electrode.

The effect of a semiconducting coating to grade the stress on the surface of the pressboard was tried in a few experiments. The coating was a carbon impregnated paper 4 inches in diameter and 0.0035 inch thick cemented to the pressboard surface before vacuum impregnation. This had slight effect on the impulse strength in the case of oil with electrode configuration C in Table II. But, in the case of Askarel-impregnated pressboard, it had a very pronounced beneficial effect when it was under the positive electrode alone and even more beneficial effect when it was placed under both electrodes. It had no effect, or, in a few cases, a deleterious effect when placed under the negative electrode alone.

With electrode configuration C with Askarel-impregnated pressboard, the impulse strength was raised to 179 kv with positive polarity when the semiconducting paper was under the top electrode. When it was at the bottom electrode alone, the impulse strength was only 124 kv with positive polarity on the top electrode. When the coating was placed under both electrodes of configuration C, the impulse dielectric strength was 194 kv average with positive polarity on the top electrode. The optimum resistivity of the several semiconducting coatings tried was about 1 megohm per square.

These tests appear to confirm again that the breakdown is being initiated in the liquid space around the positive electrode.

Doctor Sommerman[9] in some independent work has investigated the corona-like discharge which occurs in the liquid at the electrode edges in impulse tests such as are done here. With a photographic film under the electrode, he has been able to record their appearance and inception voltage and he notes that the discharge in mineral oil begins at a voltage considerably below the puncture voltage, but with Askarel it begins only slightly below the puncture voltage. This is what might be expected from the stress distribution due to the lower dielectric constant of the mineral oil compared to the pressboard. He also has suggested that the higher impulse strength of the mineral oil is due to a grading of the stress at the pressboard surface due to this corona-like discharge. That mechanism is possible, but it is not consistent with the observation that the low dielectric constant hexachlorbutadiene and hexachlorpropylene liquids had very low impulse strengths in pressboard although their 60-cycle liquid dielectric strengths were comparable with mineral oil.

It was thought that the Askarel liquid itself might show a difference in impulse strength with different polarities on an unsymmetrical electrode system, but the difference was small, if any. A 1/4-inch diameter square edge rod was tested 1/8 inch spaced from a much larger plane electrode. The impulse strength was 78 kv with a standard deviation of 1.5 with the rod positive and 74.5 with a standard deviation of 4 with the rod negative.

With the same electrode system, mineral oil had an impulse strength of 70 kv with a standard deviation of 2 with the rod positive and 77 kv with a standard deviation of 7 with the rod negative. Thus, mineral oil and Askarel liquids alone are very comparable in impulse strength. With this electrode system, the impulse ratios averaged 1.66 for the Askarel and 1.73 for mineral oil.

Summary

From the impulse dielectric strength tests described in this paper, we can conclude that a low impulse strength seems to be peculiarly associated with halogenated liquid compounds impregnated into pressboard, but this does not include the perfluorinated liquid compounds which have impulse strengths comparable with mineral oil.

It is re-emphasized that the impulse dielectric strengths of Askarel and mineral oil–impregnated pressboard are about the same as their dielectric strengths with a single cycle of 60-cycle alternating current. With the addition of multiple cycles of 60-cycle alternating current voltage, the dielectric strength of mineral oil–impregnated pressboard decreases considerably, while the dielectric strength of Askarel-impregnated pressboard decreases only a little. This effect may be due to a greater corona intensity in the case of mineral oil.

The tests with different polarities of voltage on unsymmetrical electrodes indicate that the breakdown of Askarel-impregnated pressboard is initiated in the liquid space at the edge of the positive electrode.

References

1. Dielectric Strength of Transformer Insulation, **P. L. Bellaschi, W. L. Teague.** *Electrical Engineering* (*AIEE Transactions*), volume 56, January 1937, pages 164–71.

2. Impulse Voltage Strength of Dielectrics, **J. C. Dowell, C. M. Foust.** *General Electric Review* (Schenectady, N. Y.), volume 40, 1937, pages 141–52.

3. Breakdown Curve for Solid Insulation, **V. M. Montsinger.** *Electrical Engineering* (*AIEE Transactions*), volume 54, December 1935, pages 1300–01.

4. Some Impulse-Voltage Breakdown Tests on Oil-Treated Paper-Insulated Cables, **C. M. Foust, J. A. Scott.** *Electrical Engineering* (*AIEE Transactions*), volume 59, July 1940, pages 389–91.

5. Impulse Strength of Cable Insulation, **E. W. Davis, W. N. Eddy.** *Electrical Engineering* (*AIEE Transactions*), volume 59, July 1940, pages 394–99.

6. Surge-Voltage Breakdown Characteristics for Electrical Gaps in Oil, **Royal W. Sorensen.** *Electrical Engineering* (*AIEE Transactions*), volume 59, February 1940, pages 78–81.

7. Discharge Lag in Solid Insulation, **R. Strigel.** *Wissenschaftliche Veroeffentlichungen aus den Siemens-Werken* (Berlin, Germany), volume 18, 1939, pages 101–19; Discharge Delay in Oil in Uniform Fields, **R. Strigel.** Volume 16, 1937, pages 38–50; The Time Lag in Breakdown of Oil in Non-Uniform Fields, **R. Strigel.** Volume 17, 1938, pages 38–47.

8. Impulse and 60 Cycle Strength of Air, **P. L. Bellaschi, W. L. Teague.** *Electrical Engineering* (*AIEE Transactions*), volume 53, December 1934, pages 1638–45.

9. Edge Phenomena in Impulse Breakdown Tests on Solid Insulation, **E. L. C. Larson, G. M. L. Sommerman.** *Annual Report*, Conference on Electrical Insulation, National Research Council (Washington, D. C.), 1950, pages 69–70.

Discussion

G. M. L. Sommerman (Battelle Memorial Institute, Columbus, Ohio): The suggestion that the higher impulse strength of mineral-oil-impregnated pressboard might be caused by a grading of the stress at the pressboard surface from a corona-like discharge in the oil was made by the writer while working with T. W. Dakin and C. N. Works in 1948. The basis for the suggestion was the observation that almost none of the failures in oil-immersed samples occurred at or beyond the electrode edge, while all of the failures in the Askarel-immersed samples did occur at or beyond the electrode edge.

In order to test this hypothesis, work was undertaken early in 1950 at Northwestern University by E. L. C. Larson and the writer.[1] Much of this work centered on trying to detect impulse ionization at the electrode edges in the liquid immersant by the use of photographic paper under and beyond the electrodes. The results did show that the oil immersant ionized rather uniformly around the electrodes at voltages considerably below the breakdown voltage. Also, the ionization voltage was higher for Askarel-immersed samples and was followed almost immediately by breakdown of the pressboard. With Askarel, the ionization was not uniform around the electrode but was concentrated at one or two points (at random) which, if the voltage were raised slightly, led to complete puncture. This seemed to confirm the hypothesis of a beneficial effect of ionization in the oil immersant.

Other breakdown tests did not confirm the hypothesis. Both Dakin and Works and Larson and Sommerman performed tests with oil-immersed Askarel-impregnated pressboard which gave relatively low breakdown results. The tests further showed that, in this case also, the oil was ionized rather uniformly around the electrodes. This indicates that the Askarel-impregnated pressboard does not resist well the effects of ionization in immersants. On the other hand, similar experiments with Askarel-immersed oil-impregnated pressboard gave higher breakdown results, although accompanied by Askarel-type ionization patterns in the immersant. Worthy of note here are the authors' higher breakdown results when the pressboard was preimpregnated with resin, even though the Askarel immersant retained.

Thus, it appears that the well-known benefits of immersing insulated structures in Askarel need not be accompanied by a serious reduction in impulse-breakdown strength, provided that Askarel is excluded from fibrous insulation exposed to ionization in the immersant.

The places where impulse ionization occurs in a structure may be found by the photographic-paper method.

Reference

1. See reference 9 of the paper.

T. W. Dakin: The pressboard breakdown tests referred to by Dr. Sommerman were not reported because there was some doubt about the technique wherein a pressboard sheet impregnated with one liquid is then immersed in another. Intra diffusion of the two liquids in both directions would begin immediately and continue to an undetermined extent before breakdown occurred. In the experiments to which Dr. Sommerman referred, the time of immersion before breakdown was short (about 30 minutes), but I will still think the mixing of the two liquids might be sufficient to throw some question on the result. Therefore, they cannot be considered as a positive proof that the corona on the surface does not effect the higher impulse strength of oil-impregnated and oil-immersed pressboard.

It is unfortunate that we are dealing here with a system in which it is so difficult to avoid with certainty the "electrode" and surface discharge effects in breakdown which usually so greatly reduce the breakdown strength and complicate the breakdown mechanism. It would be desirable to have tests made under conditions wherein a uniform field was produced in pressboard with no external discharge occurring before breakdown of the pressboard. This condition was approached but probably not quite reached with the semiconducting paper under the electrode, with the notable result that comparatively much higher impulse dielectric strengths were obtained with the Askarel-impregnated pressboard.

A Pulse Method for Supplying High-Voltage Power for Electrostatic Precipitation

H. J. WHITE
MEMBER AIEE

IT IS perhaps not surprising that the fundamentally important advances in the electrostatic method of separating suspended particles from gases have resulted primarily from improved methods of supplying high-voltage power. Both theory and experience bear out the great importance of the methods and apparatus used to supply electric power for the corona discharge in the precipitator, which in turn produces the electrical forces on the particles and causes their collection at a suitable electrode.

Indeed, electrical precipitation as a commercially important process dates from Cottrell's developments[1] in the period 1905 to 1910 in which he applied the then newly developed a-c transformer and synchronous mechanical rectifier to supply the high-voltage, unidirectional current required. Earlier attempts by Sir Oliver Lodge[2] to apply the process commercially in 1885 failed chiefly because of the lack of suitable high-voltage equipment. Lodge's handicap in this respect may be understood easily when it is recalled that the only source of high-voltage power available in his day was a Wimhurst electrostatic machine driven by a steam engine. Today, as in 1905, the high-voltage energizing method and equipment remains in large part the secret of successful electrostatic precipitation, and the ultimate performance and success in many cases depends more upon establishing and maintaining the proper degree of electric energization than upon any other factor.

Much recent research in the electrostatic collection of particles has been directed toward improving and stabilizing electric energization, both in connection with conventional rectifier equipment and in the development of entirely new methods of supplying electric energy to precipitators. Conventional methods of supplying high-voltage power to precipitators stem directly from Cottrell's pioneer work and essentially comprise transformer and rectifier equipment, usually with unfiltered output.

The purpose of this paper is to describe a fundamentally new system of supplying high-voltage power to Cottrell-type precipitators. The new system is based on the use of pulse principles and techniques, and therefore may be designated conveniently as "pulse energization." Basic concepts of the system may be stated as follows. In electric precipitators maximum particle collection is attained when the electrical forces acting on the particles are made as large as possible. This in turn requires maximum charges on the particles and maximum precipitation fields. The theory of particle charging[3] shows that large particle charges can be attained only by applying very high peak voltages to the precipitators, while rapid collection of the charges requires high time-average values of the voltage.

These factors are brought out by the fundamental precipitation equation[4]

$$\eta = 1 - \epsilon^{-\frac{A}{V}w} \tag{1}$$

where

η = precipitator collection efficiency
A = effective area of collecting electrodes
V = gas flow rate through precipitator
w = drift velocity of charged particles in precipitator electric field

The drift velocity w in turn may be expressed by

$$w = \frac{aE_0E_p}{2\pi\theta} \tag{2}$$

where

a = particle radius
E_0 = particle charging field
E_p = particle collecting field
θ = gas viscosity

The collection efficiency is seen to increase rapidly with increasing values of the factor w, which is a direct measure of the rate of collection of the particles. The sensitivity of the factor w to electrical conditions in the precipitator is shown by equation 2 wherein w increases as the product of the charging field E_0 and the collecting field E_p. These fields are determined by the precipitator voltage and current characteristics, and although strongly dependent on the wave forms used, are not subject to any simple analysis. The fact that both E_0 and E_p must be maximized simultaneously leads to a study of voltage wave form effects and of electric equipment for producing the optimum wave form.

In the complex conditions existing in a precipitator it is found that these requirements are best met by pulsating voltages and intermittent current wave forms, such as are supplied by ordinary unfiltered full-wave and half-wave rectifier sets. Direct current or steady voltage is characterized by a relatively low sparking voltage and unstable arcing conditions. Both the full-wave and half-wave voltages have higher sparking values than steady or pure direct voltage and are much more stable in operation. But valuable as the full-wave and half-wave voltages have proved to be, their characteristics are arbitrarily fixed by the 60 cycles per second, or occasionally 25 cycles per second, alternating voltage from which they are derived, and only by an unusual coincidence will they match precipitator requirements in any given case.

The pulse system, on the other hand, is designed so that both the duration and frequency of the current pulses supplied to the precipitator are subject to precise adjustment and control, which then makes it possible to supply optimum electric energization and thereby to insure maximum precipitator efficiency. This system essentially comprises a high-voltage, high-power pulse generator capable of supplying high-voltage pulses of the order of 100 microseconds' duration at a frequency of several hundred pulses per second, which in turn may be commutated to as many as four or even six precipitator sections. The method provides the following basic advantages as compared with conventional rectifier energization methods:

1. Higher peak voltages and increased precipitator efficiency through on-the-spot adjustment of voltage, frequency, and wave shape to fit individual precipitators.
2. Current- and power-limiting action during precipitator spark-over are inherent because of the stored-energy-type pulse circuit used.
3. Greater degree of precipitator high-tension electrode sectionalization is practicable.
4. Higher over-all electrical efficiency.

Paper 52-230, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23-27, 1952. Manuscript submitted March 28, 1952; made available for printing May 14, 1952.

H. J. White is with the Research Corporation, Bound Brook, N. J.

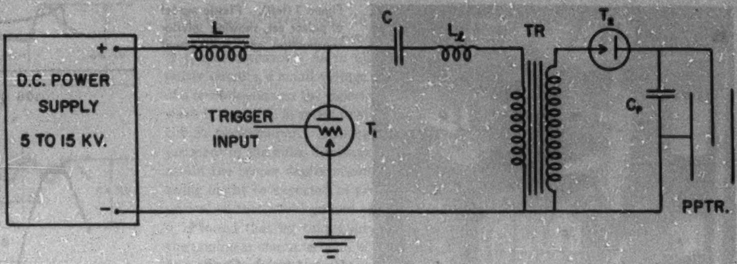


Figure 1. Fundamental circuit of pulse generator for electrical precipitation. L = charging choke, C = pulse capacitor, TR = pulse transformer, T_1 = thyratron pulse tube, T_2 = output diode, L_l = pulse inductance

Fundamental Circuit and Equations

In some respects the new pulser equipment is similar to that developed for use with high-power microwave radar equipment used during the war. The basic circuit is that of the line-type pulser,[5] in which energy is accumulated in a high-voltage capacitor over relatively long periods and then discharged very rapidly into the load. This cycle of slow charge and rapid discharge of the capacitor is repeated periodically so that the output consists of a series of uniformly spaced, high-power pulses.

The fundamental circuit is illustrated in Figure 1. In this circuit the pulse capacitor C is periodically charged from the high-voltage d-c power supply through the charging choke L and periodically discharged into the primary of the pulse transformer by means of a suitably triggered thyratron pulse tube T_1. The pulse transformer in turn raises the voltage level to that required by the precipitator, and supplies a series of short-duration current pulses to the precipitator through the hold-off diode T_2. The purpose of this output diode is to prevent power from flowing back from the precipitator through the secondary of the pulse transformer after the end of the pulse. Electric energy stored in the capacitance of the precipitator during the pulse maintains the corona current during the period between pulses. Thus, in operation the precipitator voltage rises rapidly to a peak during the pulse discharge and then slowly decays during the period between pulses, the voltage somewhat resembling a saw-tooth wave shape in form.

It may be shown easily that the circuit equations for the charging of pulse capacitor C from the d-c power supply are as follows[6]

$$V_c(t) = V_{dc} + [V_c(0) - V_{dc}] \cos \frac{t}{\sqrt{LC}} \tag{3}$$

$$i_c(t) = [V_{dc} - V_c(0)] \sqrt{\frac{C}{L}} \sin \frac{t}{\sqrt{LC}} \tag{4}$$

where

$V_c(t)$ = instantaneous value of pulse capacitor voltage
$i_c(t)$ = instantaneous value of pulse capacitor charging current
$V_c(0)$ = initial voltage on pulse capacitor
V_{dc} = d-c power supply voltage
L = charging choke inductance
C = pulse capacitor capacitance

Resonant charging is obtained when C is discharged at the time $t = T = \pi\sqrt{LC}$, which in the case of zero residual voltage on C gives a pulse capacitor voltage equal to $2V_{dc}$ at the instant of discharging. Thus, the resonant charging circuit effectively doubles the voltage of the d-c power supply and provides very high efficiency charging for pulse capacitor C.

The equations for the pulse discharge phase may be derived by suitable analysis of the equivalent circuit and are as follows

$$V_p(\tau) = V_p(0) + \frac{C}{nC_p}[V_c(T) - V_c(0)] \tag{5}$$

$$i_p(t) = \frac{1/2\sqrt{\frac{C}{L_l}}[V_c(T) - V_c(0)]}{\sqrt{1 + \frac{C}{n^2C_p}}} \sin \pi \frac{t}{\tau} \tag{6}$$

where

t = time measured from start of pulse discharge
τ = pulse duration
$V_p(0)$ = precipitator voltage at beginning of pulse
$V_p(\tau)$ = precipitator voltage at end of pulse
$V_c(T)$ = pulse capacitor voltage at start of pulse
n = turns ratio of pulse transformer
C_p = capacitance of precipitator high-tension circuit
L_l = leakage inductance of pulse transformer

By the use of these equations it is possible to make a complete analysis of the operation of the pulse circuit when losses are neglected. More refined equations in which losses are included permit closer calculations to be made but are omitted here because their more complicated forms tend to obscure the fundamentals involved.

Pulse Generator Design for Precipitator Applications

Cottrell-type precipitators generally require corona voltages in the range of 30 to

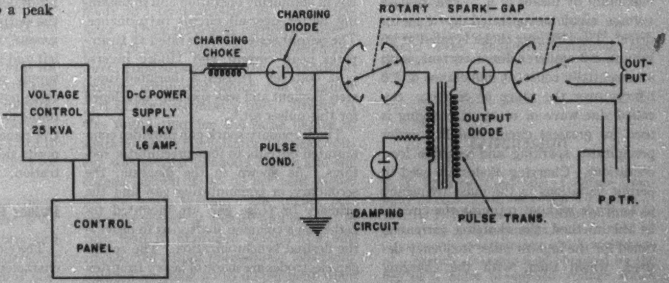


Figure 2. Schematic diagram for experimental pulser set

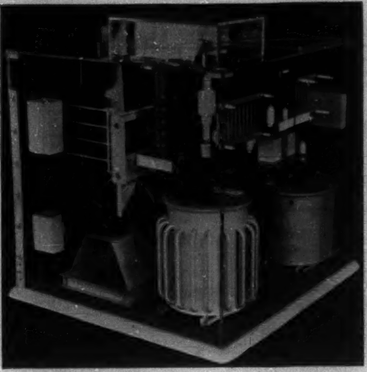


Figure 3 (left). Plastic model of pulser set, showing interior construction and arrangement

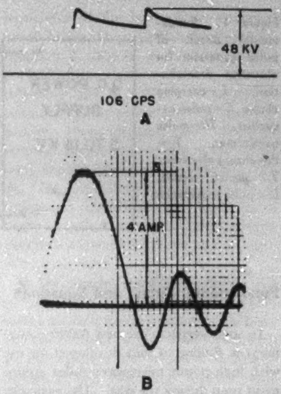


Figure 4 (right). Oscillograms of typical precipitator voltage and current wave forms for pulser operation

(A) Voltage wave form
(B) Current wave form

100 kv, and corona currents up to several hundred milliamperes.[7] In the case of the pulser equipment, multiple high-voltage outputs are used in order to provide commutated power for several precipitator sections. The reason for the latter provision is that much higher precipitation efficiencies generally are obtained by using a number of separate high-tension electrode sections which are isolated electrically from one another.

From the results obtained in the laboratory and field on several experimental pulser sets the following specifications were arrived at for a commercial unit:

Peak output voltage, 70 kv
Average or d-c output current, 300 milliamperes
Power output, 15 kw
Pulse frequency range, 240 to 480 cycles per second
Pulse duration, 100 to 150 microseconds
Number of commutated outputs, 4

A schematic diagram of the commercial pulser set is shown in Figure 2. In this circuit, high-voltage d-c power is supplied by a 14-kv 1.6-ampere 3-phase tube rectifier set. The voltage output of the rectifier is continuously variable from zero to maximum by means of a 25-kva 3-phase voltage regulator operated by a motor drive. The charging choke is rated at 30 kv and 2.2 amperes direct current, with an essentially constant inductance of 3.0 henrys over the range of current. So-called sine wave or resonant charging is used for greatest circuit stability under precipitator sparking and variable load conditions. Charging diodes are used to permit variations in the pulse frequency to be made without retuning the circuit. In this method, the charging current is timed for the highest pulse frequency desired, which then, with the charging diode, permits lower pulse frequencies to be used without retiming.

Because of the very high voltage and current requirements, as well as the long life necessary, a rotary spark-gap pulse switch is used instead of a thyratron tube. The pulse capacitor is comprised of a number of separate units connected in parallel, permitting step adjustment of the capacitance between about 0.1 and 0.2 microfarad. In operation, the required value of the pulse capacitance is found by trial to correspond to the power requirements of the precipitator with which the pulser is used. The pulse inductance is incorporated in the pulse transformer in the form of leakage reactance, the correct value being obtained by proper design of the leakage reactance paths in the transformer.

The pulse transformer has a step-up ratio of 5.6 to 1 and is designed to handle conservatively the full power output of the pulser without the overheating which is sometimes associated with pulse equipment. The damping circuit is used across a portion of the pulse transformer primary to absorb the residual transients which occur after each pulse. This is advantageous in stabilizing operation and in obtaining better over-all circuit performance. The output series diode is rated at 10 amperes peak current and 110-kv inverse peak voltage. It uses a thoriated tungsten filament and was specially developed for this pulser set.[8]

The secondary spark gap provides commutated outputs to four precipitator sections. As shown in the diagram, the secondary or commutating gap and the primary or pulse gap are mounted together on a common shaft so as to provide the desired synchronization. The spark-gap electrodes are made of heavy tungsten rods and have an expected service life of 6 to 12 months or longer.

One of the cardinal advantages of the line-type pulser for precipitator service is that inherent protection is provided against the sparking which usually occurs in precipitator operation. The protection is intrinsic and fundamental to the pulse circuit itself. Thus, on each pulse the energy-storage capacitor C is completely discharged, and if a spark happens to occur in the precipitator during a pulse, the current will momentarily be increased somewhat, but no more energy can be supplied than on a normal cycle. This situation is in sharp contrast to that for rectifier equipment, for which it is always necessary to use some form of protective series resistance or other current-limiting means to prevent catastrophic power line surges when the precipitator sparks. Such protective equipment is wasteful of power and in its more complicated forms tends to compromise the reliability of the entire equipment.

The general layout and assembly of the pulser set are shown in Figure 3, which shows a scale model constructed of transparent plastic in order to bring out the interior details. The pulse transformer and the output diode are shown in the foreground, with the rotary spark gap on the left and the charging choke and d-c power supply on the right. The four output cable potheads are seen near the top and to the rear of the output diode. The control panel and access door are on the opposite side and are not visible in the illustration.

Pulser Performance

The electrical operation of the pulser is characterized by a desirably high over-all

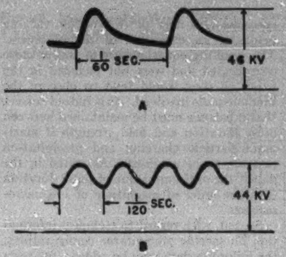


Figure 5. Typical wave forms for half-wave and full-wave rectified voltages taken in the field on a fly ash precipitator

(A) Half-wave
(B) Full-wave

power efficiency of the order of 70 per cent, and by flexibility of power output, pulse frequency, and pulse duration. Pulse frequency is easily varied by means of a variable speed motor on the rotary spark gap, while the pulse duration may be varied by changing the pulse capacitance, by adding capacitance in shunt with the precipitator, or by adding inductance in series with the primary of the pulse transformer. In actual operation, these adjustments are individually made for each precipitator so as to provide highest collection efficiency.

Oscillograms of typical precipitator voltage and current wave forms for pulser operation are shown in Figure 4. The voltage oscillogram was recorded during tests made in the field on a large fly ash precipitator, while the current oscillogram was made in the laboratory in connection with development studies of the equipment. The tail oscillations in the current pulse are caused by decay of the stored electric and magnetic energy in the pulse transformer, but are not transmitted to the precipitator because of the blocking action of the output diode. In practice, these oscillations are heavily damped by means of the damping diode and resistor which are placed across the primary of the pulse transformer.

Figure 5 shows typical wave forms for half-wave and full-wave rectified voltages taken in the field on a fly ash precipitator. Comparison of these with the pulse voltage wave form of Figure 4(A) will serve to indicate the differences in waveform for the several types of energization.

Operation of the pulse capacitor charging circuit is illustrated in Figure 6 by oscillograms of the charging voltage and current wave forms. The peak operating voltage used on the pulse capacitor normally ranges between about 15 kv and 25 kv, depending on the power requirements of the precipitator. As in all line-type pulser circuits, a small voltage overswing of a few kilovolts on the capacitor voltage wave is desirable in order to stabilize the circuit performance. Although the adjustment of the pulse discharge circuit to attain the proper degree of voltage overswing might be expected to prove rather critical on theoretical grounds, in practice it is found that by taking advantage of the nonlinear characteristics of the transformer core, the adjustment is not at all critical and may be maintained over wide ranges of conditions.

The effective over-all voltage step-up ratio of the pulse circuit, measured from rectifier output to pulse transformer secondary output, is approximately equal to the step-up ratio n of the pulse transformer. A voltage gain of about two is obtained in the resonant charging circuit, and a further gain of about $n/2$ in the pulse transformer itself.

Precipitator sparking is reflected in the pulse circuit by a momentary increase in the pulse discharge current, but with proper adjustment this produces little or no effect in the charging or d-c power supply circuits. Thus a precipitator spark will be indicated by only a small, momentary variation in the rectifier current and usually is imperceptible in the a-c input circuit. In fact, it is possible to short-circuit completely the output of the pulser without producing a noticeable flicker in the input circuit. In practice this means that protection against accidental short circuits in the precipitator is best provided by means of a d-c relay placed in the secondary circuit of the pulse transformer. Also, if it is desired to take a given section of the precipitator out of service, it is only necessary to disconnect the corresponding input terminal to the rotary spark-gap switch.

Extensive field tests of the effectiveness of the pulse system of energization on precipitator collection efficiency have been made on a representative fly ash precipitator. These tests were carried out over a period of about a year, and included many direct comparisons with the conventional rectifier equipment which was a regular part of the precipitator unit. Briefly, it was found that for the same degree of sparking in the precipitator, the peak voltage was several kilovolts higher and the corona current 20 to 35 per cent higher for the pulse system. The higher permissible operating voltage and current in turn lead to improved precipitator efficiency, the actual precipitator loss being reduced by some 50 to 60 per cent when using the pulse system as compared with the ordinary rectifier system. The reliability of the pulse method and equipment has been satisfactorily tested by continuous operation of an experimental unit on a large fly ash precipitator for a period of somewhat over a year.

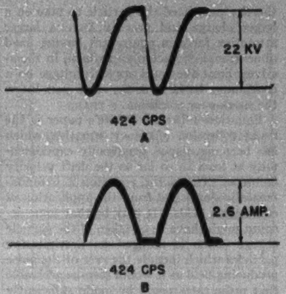


Figure 6. Voltage and current wave forms for charging of pulse capacitor

(A) Voltage wave form
(B) Current wave form

References

1. Problems in Smoke, Fume and Dust Abatement, F. G. Cottrell. Smithsonian Report for 1913, U. S. Government Printing Office (Washington, D. C.), pages 653–85.

2. The Electrical Disposition of Dust and Smoke with Special Reference to the Collection of Metallic Fumes and to a Possible Purification of the Atmospheres, O. J. Lodge. *Journal*, Society of Chemical Industry (London England), volume 5, 1886, pages 572–76.

3. Particle Charging in Electrostatic Precipitation, H. J. White. *AIEE Transactions*, volume 70, part II, 1951, pages 1186–91.

4. Electrostatic Collection of Fly Ash, H. J. White, L. M. Roberts, C. W. Hedberg. *Mechanical Engineering*, American Society of Mechanical Engineers (New York, N. Y.), November 1950, pages 873–80.

5. Pulse Generators (book), edited by G. N. Glasoe, J. V. Lebacqz. M.I.T. Radiation Laboratory Series, volume 5, McGraw-Hill Book Company, Inc., New York, N. Y., 1948, page 8.

6. *Ibid.*, chapter 9.

7. The Role of Corona Discharge in the Electrical Precipitation Process, H. J. White. *Electrical Engineering*, volume 71, January 1952, pages 67–73.

8. High Voltage Rectifier Tubes for Extremely High Voltage and Power, T. H. Rogers. *Cathode Press*, Machlett Laboratories, Inc., (Springdale, Connecticut), Summer Issue 1949, page 2.

Discussion

E. R. Thomas (Consolidated Edison Company of New York, Inc., New York, N.Y.): Mr. H. J. White is to be complimented on presenting a new concept in a method of electrostatic precipitation which makes use of a particle charging field obtained by the use of a pulse of applied voltage which thereby permitted the particle to take on a larger charge and thereby have a larger attractive force in equipment having fixed dimensional characteristics which, in themselves, limit the total applied voltage on a continuous basis to some value just below the spark-over condition.

Equation 1 in Mr. White's paper is the usual collection efficiency equation which has been developed principally experimentally at least in so far as the drift velocity factor *w*. The factor, equation 2, is fundamentally developed from the application of continuous fields. It may be that some account might have to be taken of the particle charge which is induced on precipitation particles which include the rate of the pulse producing field as it would seem conceivable that pulse rates could be made sufficiently high so that a maximum particle charge for a given peak pulse field would not be obtained.

I would like to ask the author of this paper if he has considered expressing the evaluation of the drift velocity factor in terms of peak applied voltages and pulse rates.

Since the author has made extensive field tests on this pulse system of precipitation, I believe that the inclusion of actual technical data on specific precipitator configurations using both the continuous type of electrification and the pulse type of electrification would be of extreme interest and might be included with the paper.

H. J. White: First, with regard to the theoretical equations, it should be emphasized that these are based on somewhat idealized conditions which are never fully realized in practice. For this reason, and also because the fundamental physical quantities in equation 2 are seldom known with any degree of accuracy, it is usually necessary to evaluate the drift velocity factor *w* from direct experimental data rather than by theoretical deduction. The fact that equation 2 is fundamentally developed for continuous fields is not actually a limitation of the theory because a similar equation also may be derived for varying fields by an extension of the argument. The resulting equations are of course more complicated and were not included in the paper as they would tend to obscure the fundamentals involved. It is indeed correct that a balance must be maintained between pulse duration and field strength if maximum particle charging and precipitation rate are to be attained. As stated in the paper, the optimum value of pulse duration is of the order of magnitude of 100 microseconds.

Second, with regard to technical information on specific precipitator configurations, the following data may be added: the tests were made on two separate precipitators treating approximately 250,000 cubic feet per minute and 500,000 cubic feet per minute respectively, both having rated collection efficiencies of 95 per cent. These precipitators were of the plate type and were operated at gas velocities of about 10 feet per second. The duct spacing was 8 inches and the duct length 18 feet.

Electrical and Physical Properties of IN-420: A New Chlorinated Liquid Dielectric

A. J. WARNER
MEMBER AIEE

Synopsis: A family of new chlorinated liquid dielectrics has been developed. One member of this series, dichloro-1,3,3-trimethyl-1-phenylindan (designated as IN-420), appears particularly suitable for use as a capacitor impregnant. Values of dissipation factor and d-c resistivity are comparable to or somewhat better than existing commercial chlorinated biphenyls. Its dielectric constant and dielectric strength are somewhat superior to those of the chlorinated biphenyls.

In life test studies in which capacitors impregnated with IN-420 were compared with capacitors containing stabilized and unstabilized chlorinated biphenyls, a marked superiority was shown in life test characteristics and particularly at temperatures as high as 125 degrees centigrade.

THE USE of chlorinated hydrocarbons in electric apparatus and components has increased steadily in the past few years, particularly in applications such as capacitors where the use of high dielectric constant materials is advantageous.[1-3] Continued research and development has produced materials useful over a wide temperature range and possessing good performance characteristics under prolonged electrical stress. Typical of such materials are the chlorinated biphenyls (Aroclors) and the chlorinated naphthalenes (Halowaxes). It is the purpose of this paper to introduce a new family of chlorinated hydrocarbons, although details will be given for just one member of the family at this time.

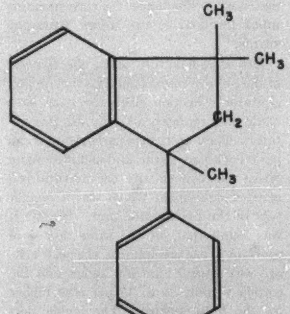


Figure 1. Structural formula of 1,3,3-trimethyl-1-phenylindan

Chlorinated Indans

For many years the Federal Telecommunication Laboratories has been interested in the substituted phenylindan derivatives and particularly in the compound having the structural formula shown in Figure 1.

This material can be obtained in good yields, and economically, from available

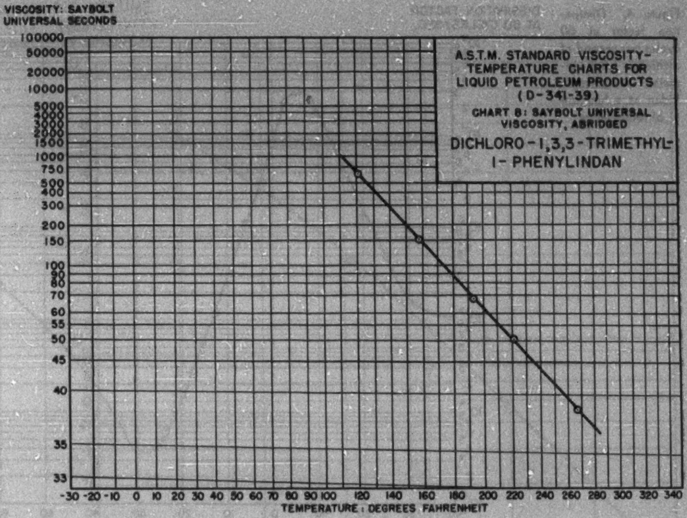
Table I. Physical Properties of IN-420 and Aroclor 1254

Property	In-420	Aroclor 1254
Refractive index at 25 degrees centigrade	1.5838	1.6380
Specific gravity at 65 degrees per 15.5 degrees	1.141	1.502
Pour point, degrees centigrade	8	8
Acidity, milligrams NaOH per gram	0.001	0.001
Free inorganic chlorides, parts per million	0.1	0.05
Distillation range, degrees centigrade		
First drop	340	...
10 per cent	343	370
40 per cent	347	376
90 per cent	356.6	387
Flash point, degrees centigrade	185	...
Fire point, degrees centigrade	250	...

Paper 52-227, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23-27, 1952. Manuscript submitted February 25, 1952; made available for printing May 26, 1952.

A. J. WARNER is with the Federal Telecommunication Laboratories, Inc., Nutley, N. J.

Figure 2. Viscosity of dichloro-1,3,3-trimethyl-1-phenylindan as a function of temperature



raw materials and by relatively simple processes. It is a crystalline material having a melting point of 52 degrees centigrade, a boiling point of 145 degrees centigrade at a pressure of 4 millimeters, and a refractive index at 20 degrees centigrade of 1.5660. At its melting point it has a dielectric constant of 2.55 and a dissipation factor of 0.001_4. Despite the fact that it is not possible to construct a model of this compound using conventional Fisher-Hirschfelder building blocks, which is normally taken as indicating that such a structure, if possible, would be highly unstable, the substituted indan has been found to possess a high order of stability.

This high stability encouraged the preparation of the chlorinated derivatives, and it has been found possible to prepare the full range of products up to and including the nonachloro derivative, ranging from a mobile light-colored liquid to high-melting brittle resins. Although data have been accumulated on the whole range of products obtained, the particular properties of the dichloro derivative were such as to offer immediate application possibilities, and these properties have been studied most extensively.

Dichloro-1,3,3-trimethyl-1-phenylindan, to which the trivial name IN-420 has been given, is a clear, lightly yellow-colored, medium viscosity material having a bluish fluorescence. Representative physical data, in comparison with pentachlorobiphenyl (Aroclor 1254), are shown in Table I. The data for IN-420 and Aroclor 1254 were determined in our laboratories; those obtained for the latter impregnant are in general accord with the information contained in the trade literature.[4]

The viscosity of IN-420 as a function of temperature is given in Figure 2.

The electrical properties of this material have been extensively investigated using a rhodium-plated, guarded test cell. The 60-cycle dielectric constant and dissipation factor measurements were made on a modified 60-cycle Schering bridge operating at 3,000 volts rms, and with a dielectric voltage stress of 37.5 volts per mil. D-c volume-resistivity measurements were made with a General Radio type 544-B Megohm bridge. Measurements of dissipation factor and dielectric constant at frequencies from 400 to 100,000 cycles per second were made on a special audio-frequency test set. The temperature of the test cell was maintained within 0.2 degree centigrade for a 45-minute period previous to each measurement.

The curves of dielectric constant and dissipation factor at 60 cycles per second versus temperature, are plotted in Figures 3 and 4. The dissipation factor above 30 degrees centigrade is purely ionic, increasing logarithmically with increasing temperature. Below 30 degrees centigrade, the dissipation factor increases, going through a maximum at −6 de-

Figure 3. Dielectric constant at 60 cycles per second of dichloro-1,3,3-trimethyl-1-phenylindan as a function of temperature

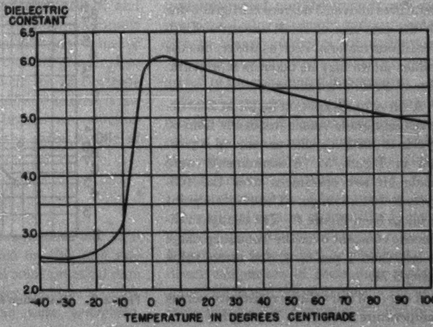
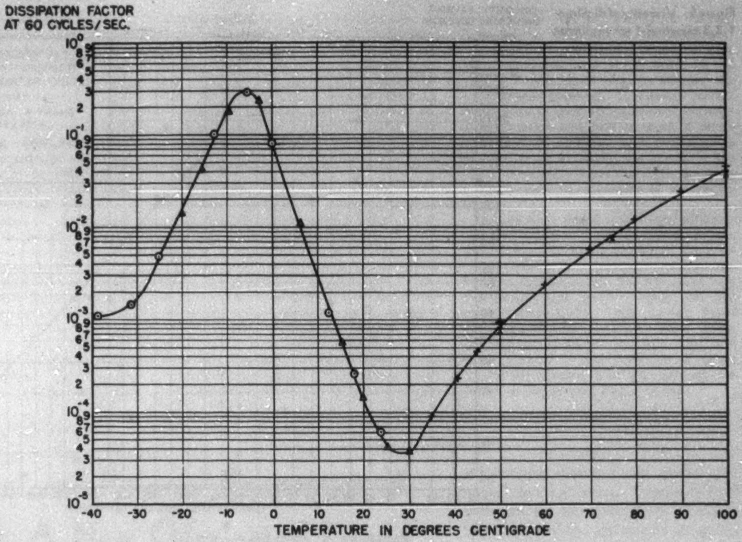


Figure 4. Dissipation factor at 60 cycles per second of dichloro-1,3,3-trimethyl-1-phenylindan as a function of temperature



grees, then decreasing. The 60-cycle dielectric constant decreases as the temperature is increased from 5 to 100 degrees centigrade with an average relative change of 0.012 per degree centigrade due primarily to decreasing density. Below 5 degrees centigrade, the dielectric constant decreases rapidly, becoming asymptotic to a value of approximately 2.58. The inflection point of this curve occurs at approximately −6 degrees centigrade. The phenomenon centering about the −6-degree point is the dipolar anomalous dispersion, the molecule dipole being in resonance with the applied 60-cycle voltage at this temperature.

Individual samples may show slight differences in dissipation factor at temperatures above 30 degrees centigrade due to the presence of small amounts of adventitious material such as water, but the values shown may be taken as representative of clean, dried material.

A curve of dissipation factor at 24.0 degrees centigrade versus frequency from 60 cycles to 100 kilocycles per second is plotted in Figure 5. Measurements were made on two specimens over the 400-cycle to 100-kc range. The 60-cycle point is taken from Figure 4. The change in dielectric constant over the frequency range is less than 1 per cent, the mean value being 5.71.

A curve of d-c volume resistivity versus temperature from 40 to 100 degrees centi-

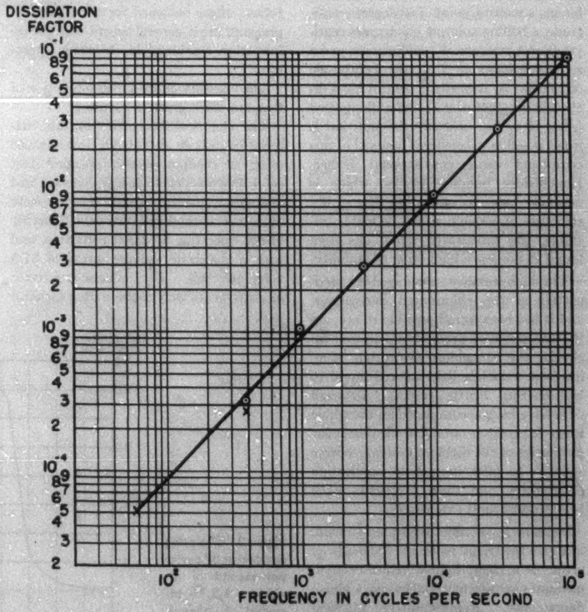
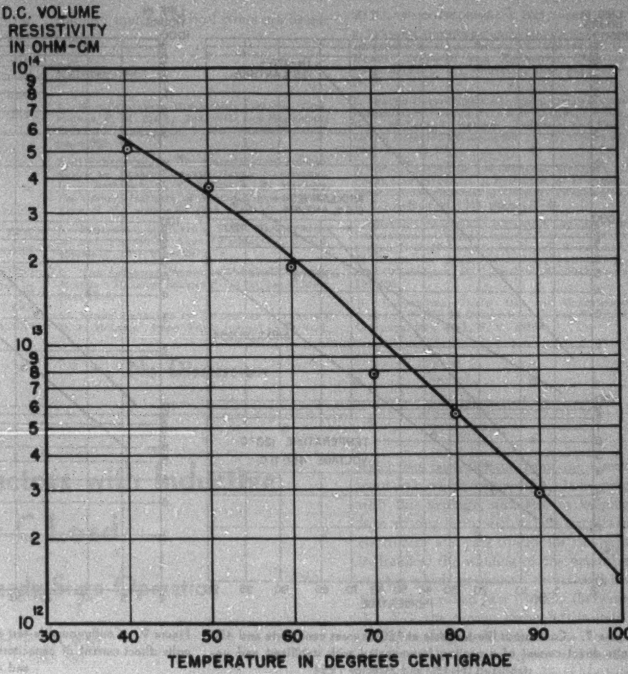


Figure 5. Dissipation factor at 24 degrees centigrade of dichloro-1,3,3-trimethyl-1-phenylindan as a function of frequency

Figure 6. D-c volume resistivity of dichloro - 1,3,3 - trimethyl - 1- phenylindan as a function of temperature



grade is plotted in Figure 6. The points plotted are the averages of measurements made on four specimens. These measurements were made after an electrification period of 3 minutes. Although a 1-minute electrification is the more common, it was found that use of a 3-minute period resulted in the obtaining of more stable values and hence in an improved reproducibility of measurements.

The corresponding data for Aroclor 1254 indicate a general similarity, and for comparison we may take data at 50 and 100 degrees centigrade as shown in Table II.

Table II. Electrical Properties of IN-420 and Aroclor 1254

Material	Temperature, Degrees Centigrade	Tan δ	ε′
IN-420	50	0.00085	5.31
	100	0.0400	4.90
Aroclor 1254	50	0.0011	4.80
	100	0.0427	4.25

The values of dissipation factor for IN-420 are slightly better than for Aroclor 1254 and the d-c resistivity values for the two materials are comparable. IN-420, however, has a somewhat higher dielectric constant of 4.90 at 100 degrees centigrade as compared with 4.25 for Aroclor 1254 at the same temperature.

The dielectric strength properties of impregnants are obviously of great importance to the user of such materials for electric apparatus and components, and comparative tests in our laboratory made on IN-420 and Aroclor 1254 gave values of greater than 45 kilovolts and greater than 35 kilovolts respectively for an electrode spacing of 100 mils. Also of interest is the maintenance of high breakdown values on repeated testing of the same sample of IN-420.

Life Test Studies

On the basis of the physical and electrical results obtained on this new material, it was felt that some practical data on its use potentialities would be in order. Therefore, capacitors were assembled, comprising 0.0005-inch thick aluminum foil separated by two thicknesses of 0.0004-inch unbleached kraft condenser tissue (Peter J. Schweitzer, Inc.), having an approximate capacity of 0.06 microfarad when impregnated. Kraft tissue was used in assembling these capacitors in view of the reported superiority of this material compared with linen paper.[5,6] For comparison purposes, Aroclor 1254 was used as an impregnant in half of the units assembled, and the greatest care was exercised during assembly and impregnation to ensure minimum contamination. In some experiments stabilizers[7,8] were added to the chlorinated impregnant.

Some of the units were put on continuous life test at 120 degrees centigrade and others at 125 degrees centigrade, while the voltages used were 450 volts direct current and 600 volts direct current. The results obtained have been plotted in a conventional manner on logarithmic probability paper, from which the life value for 50-per-cent failure can be obtained. A direct comparison of these data enables some measure of the usefulness of the new impregnant to be determined. Figure 7

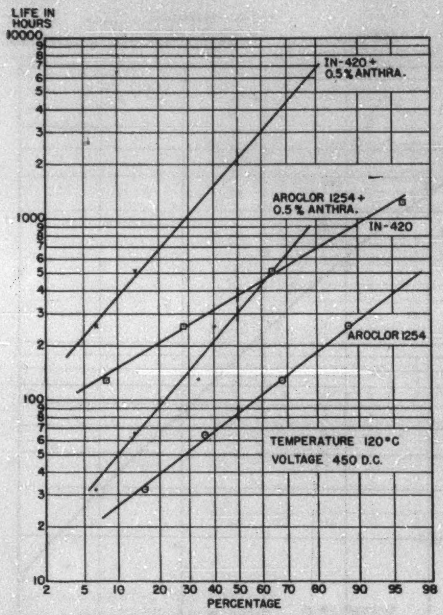


Figure 7. Continuous life-test data at 120 degrees centigrade and 450 volts direct current of capacitors impregnated with stabilized and unstabilized IN-420 and Aroclor 1254

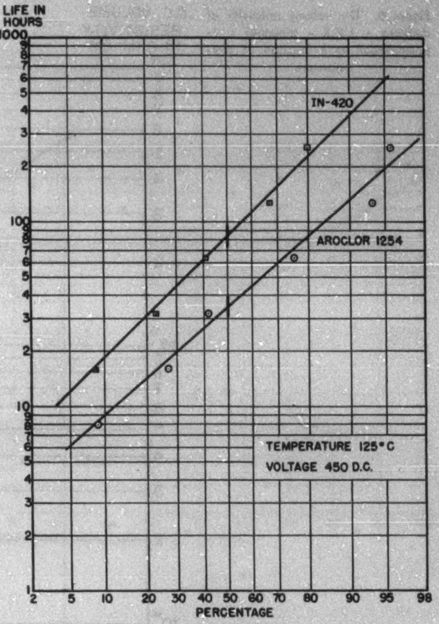


Figure 9. Continuous life-test data at 125 degrees centigrade and 450 volts direct current of capacitors impregnated with unstabilized IN-420 and Aroclor 1254

gives the data for 120 degrees centigrade and 450 volts, Figure 8 that for 120 degrees centigrade and 600 volts, and Figure 9 that for 125 degrees centigrade and 450 volts.

From Figures 7 and 8 it will be seen that the IN-420 is superior in life characteristics under the stated conditions to Aroclor 1254, even when the Aroclor is improved in a known manner by the addition of 0.5-per-cent anthraquinone. Thus, as is seen in Figure 7, the 50-per-cent failure point for plain Aroclor 1254 was 85 hours; for Aroclor 1254 plus 0.5-per-cent anthraquinone, it was 300 hours; whereas for plain IN-420 the value was 375 hours. Should 0.5-per-cent anthraquinone be added to the IN-420, then the same stabilizing influence of the additive is evidenced, the 50-per-cent failure point for this mixture being in excess of 2,000 hours. Similarly, as shown in Figure 8, the 50-per-cent failure point for capacitors impregnated with stabilized IN-420 at 600 volts direct current was 250 hours as compared with 122 hours for the stabilized Aroclor 1254.

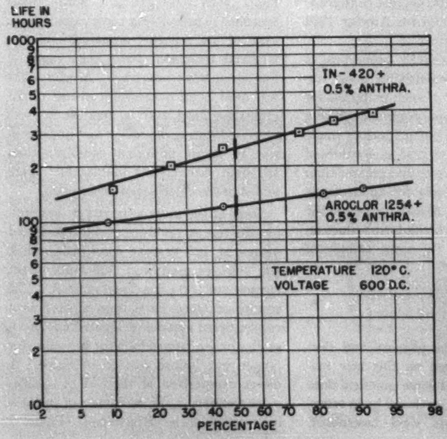


Figure 8. Continuous life-test data at 120 degrees centigrade and 600 volts direct current of capacitors impregnated with stabilized IN-420 and Aroclor 1254

At the somewhat higher temperature of 125 degrees centigrade, as shown in Figure 9, a corresponding superiority was maintained, the 50-per-cent failure points being 35 hours and 85 hours for the unstabilized Aroclor 1254 and IN-420 respectively.

The life test studies reported herein were performed under d-c test conditions because of the severer and more rapid breakdown of impregnated capacitors under such a stress.[9,10] Preliminary results reported by an outside laboratory on capacitors impregnated with IN-420 and evaluated under a-c test conditions are very promising. Further evaluation of a-c voltage applications is continuing, and these results will be reported at another time.

Conclusion

It will be seen, therefore, on the basis of the foregoing, that IN-420, a new chlorinated liquid dielectric, possesses physical and electrical characteristics indicating its usefulness particularly in capacitor applications and for use at elevated temperatures where its remarkable stability under electrical stress can be exploited.

References

1. Nonflammable Dielectric Organic Compounds, F. M. Clark. *Industrial and Engineering Chemistry* (Washington, D. C.), volume 29, 1937, pages 698–702.

2. Characteristics of Chlorinated Impregnants in Direct-Current Paper Capacitors, L. J. Berberich, C. V. Fields, R. E. Marbury. *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 33, 1945, pages 389–97.

3. Electrical Insulation, F. M. Clark. *Chemical and Engineering News* (Washington, D. C.), volume 25, 1947, pages 2976–78.

4. The Aroclors, Monsanto Chemical Company. *Monsanto Technical Bulletin P-115* (St. Louis, Mo.), June 1950.

5. Paper Problems in Telephone Apparatus, D. A. McLean. *Paper Trade Journal* (New York, N. Y.), volume 118, number 4, 1944, pages 31–34.

6. Paper Capacitors Containing Chlorinated Impregnants, D. A. McLean, L. Egerton, C. C. Houtz. *Industrial and Engineering Chemistry* (Washington, D. C.), volume 38, 1946, pages 110–16.

7. Paper Capacitors Containing Chlorinated Impregnants—Stabilization by Anthraquinone, D. A. McLean, L. Egerton. *Industrial and Engineering Chemistry* (Washington, D. C.), volume 37, 1945, pages 73–79.

8. Stabilization of Chlorinated Biphenyl in Paper Capacitors, L. J. Berberich, R. Friedman. *Industrial and Engineering Chemistry* (Washington, D. C.), volume 40, 1948, pages 117–23.

9. Paper Dielectrics Containing Chlorinated Impregnants; Deterioration in Direct-Current Fields, D. A. McLean, L. Egerton, G. T. Kohman, M. Brotherton. *Industrial and Engineering Chemistry* (Washington, D. C.), volume 34, 1942, pages 101–109.

10. Capacitors—Their Use in Electronic Circuits (book), M. Brotherton. D. Van Nostrand Company, New York, N. Y., 1946.

No Discussion

Saturable Reactors with Inductive D-C Load

Part I. Steady-State Operation

H. F. STORM
MEMBER AIEE

THE operation of saturable-reactor circuits with purely resistive loads, such as incandescent lamps, has been described in many articles.[1–6] There are, however, a great many applications especially in connection with feedback control systems where the load consists of an inductance, such as the field of an electric machine, requiring direct current. Yet the writer is not aware of any published mathematical analysis of the inductive d-c case, including effects of variations of line voltage, line frequency, and load resistance. After the completion and first submission of this paper, Mr. Milnes' paper[12] dealing, in part, with saturable reactors with inductive d-c load became known to the author.

Description of Circuit

In Figure 1, the a-c supply bus *1* energizes the windings *X1-X2* of two saturable reactors which are connected in series. A rectifier type d'Arsonval ammeter is interposed between *3* and *4* to read the average current I_G. The load consists of an impedance with the load inductance L_L and the load resistance R_L, energized from the bridge-connected load rectifier with the branches $REC_{1,2,3,4}$, which, in turn, closes the power circuit by connecting to the other pole *5* of the a-c supply.

As will be shown later in full detail, the impedance across terminals *X1*, *X2* is very large with respect to the load impedance across terminals *4*, *5* during the first part of each half-cycle of the supply voltage, whereas during the remainder of the half-cycle the reverse is true.

Because of this periodic appearance and disappearance of a large impedance in series with the load, the operation of the two saturable reactors can be likened to that of a synchronous switch, operating between supply voltage and load. This switch is open in the first part of the half-cycle, and it is closed during the remainder of the half-cycle; the phase position of the instant of closure of the synchronous switch determines the magnitude of the load current. The last sentence applies equally well to a thyratron circuit and thereby points to an arresting analogy between the two circuits. The action of the thyratron in blocking or admitting the line voltage is well described by its name which has been derived from the Greek word θυρα which means door. In keeping with this analogy, and also in keeping with existing terms such as gate circuits in electronics[7] and gate control for valves in hydraulics, the winding of the saturable reactor which produces "thyratron" action will be called gate winding (between *X1*, *X2*), the current through it is called gate current i_G, the voltage across the gate winding is the gate voltage e_G, and the resistance and reactance of the gate winding is called gate resistance R_G and gate reactance X_G respectively. Gate current i_G and load current i_L are not necessarily identical.

The phase position at which the gate current begins to flow is called the starting angle α and corresponds to the firing angle of the thyratron. The starting angle, and hence the load current, is controlled by the current flowing through the windings with the terminals *Y1*, *Y2*. These windings are called control windings, the current through it is the control current, i_C and the driving voltage, which is the independent variable, is called the control voltage E_C.

The gate windings also can be connected in parallel by connecting the odd-numbered terminal of one gate winding to the even-numbered terminal of the other gate winding. It can be shown that for a very small control circuit resistance

---

Paper 52-236, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted March 17, 1952; made available for printing May 14, 1952.

H. F. Storm is with the General Electric Company, Schenectady, N. Y.

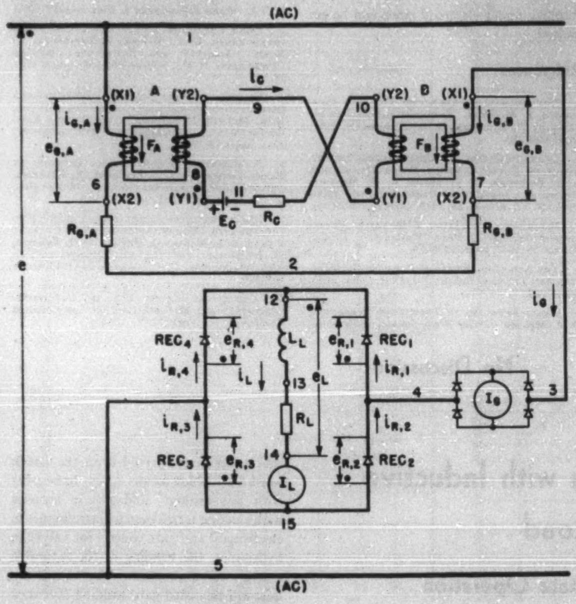


Figure 1. Schematic diagram of series-connected saturable reactor with inductive d-c load

($R_C' \ll R_L$), or for a very small gate resistance ($R_G \ll R_L$), the steady-state operation of the parallel-connected saturable reactor becomes analogous to the series-connected saturable reactor with low control-circuit resistance. All expressions derived for the latter circuit are applicable for the parallel-connected saturable reactor by replacing N_G by $(1/2)N_G$, reference 8.

The following discussion deals with the saturable reactor with so-called low control-circuit impedance. A saturable reactor belongs to the low control-impedance category, when the (even harmonic) alternating currents in the control circuit are practically the same as if encountering zero control-circuit impedance. In practice, the control-circuit impedance is low when it is a fraction of the load resistance, based on a unity turns ratio.

Saturable reactors with so-called high control circuit impedance[9] are not specially discussed because their operation in the proportional region is similar for resistive and inductive d-c load.

Exciting Interval

It will be assumed in the following that the time constant L_L/R_L is very large with respect to the duration of a half-cycle, and therefore the ripple of the load current becomes negligible. During the exciting interval ($0<\omega t<\alpha$) the gate is shut and the gate current i_G is zero; see Figure 2(E). This fact, however, does not infer that the load current i_L must also be zero; see Figure 2(D). If a load current existed at $\omega t=0$, this load current can continue to flow by circulating through the load rectifier even though the gate current is zero. With identical rectifiers, the currents in all rectifier branches are equal, and hence there is no voltage between *4* and *5*. The foregoing leads to two conclusions:

1. In contrast to the case with resistive load, gate current and load current are not necessarily the same in the case with inductive d-c load.
2. The circulating load current does not introduce a voltage difference between *4* and *5*.

The next question relates to the effect of the exciting current on the behavior of the circuit. Assume identical and constant forward resistances in all branches of the load rectifier, and let half of the load current $I_L/2$ flow through each branch. To a current i_G which tries to pass from *4* to *5* or vice versa, the circuit will look linear as long as $i_G<I_L$. Applying superposition, current i_G splits into equal parts, each half flowing through REC_1 and REC_2 respectively. The drops across these rectifiers are the same, hence no voltage difference has been introduced by current i_G between points *12* to *15*, and therefore there is no reason for current i_G, or part of it, to enter the load. Each half of i_G continues to flow through REC_3 and REC_4 respectively, finally combining to i_G at the juncture of the two rectifiers. The total resistance of the load as seen between *4* and *5* equals the forward resistance of one branch rectifier (for example, REC_1). One arrives at one more conclusion:

3. A gate current i_G which is smaller than the load current i_L can pass through the rectifier bridge without affecting the load current and it encounters only the forward resistance of one rectifier branch.

Since the exciting current of the saturable reactor is usually very small with respect to the load current, it passes through the load rectifier, causing only a very small voltage drop on the forward resistance of the rectifiers. Therefore, for all practical purposes, the excitation of units *A* and *B* will occur as though points *4* and *5* were directly connected. The core fluxes for the saturable reactor with resistive load have been calculated under precisely the same circumstances in reference 1. The results are therefore immediately available and are plotted in Figures 2(B) and (C) from Figures 3(D) and (E) of reference 1. As in Figure 2 of reference 1, the (flux-density–field intensity) (B-H) curve again will be assumed single-valued and rectangular.

Conduction Interval

During the conduction interval ($\alpha<\omega t<\pi$) the gate is open and the line voltage is acting across the points *4* and *5*, provided the voltage drops across the gate resistances $R_{G,A}$, $R_{G,B}$ and the gate current ammeter are neglected. It has been shown previously that the impedance presented by the bridge across *4* to *5* is equal to the forward resistance of one branch of the load rectifier as long as the gate current i_G is smaller than the load current i_L. Let the output voltage be defined as e_0, the line voltage during the conduction interval. Then, under the presence of the output voltage e_0, Figure 2(A), the gate current will increase sharply at $\omega t=\alpha$ until it reaches the magnitude of the load current. At this point, the current direction in two conjugate rectifier branches would reverse if it were not for the blocking action of the rectifiers which is now

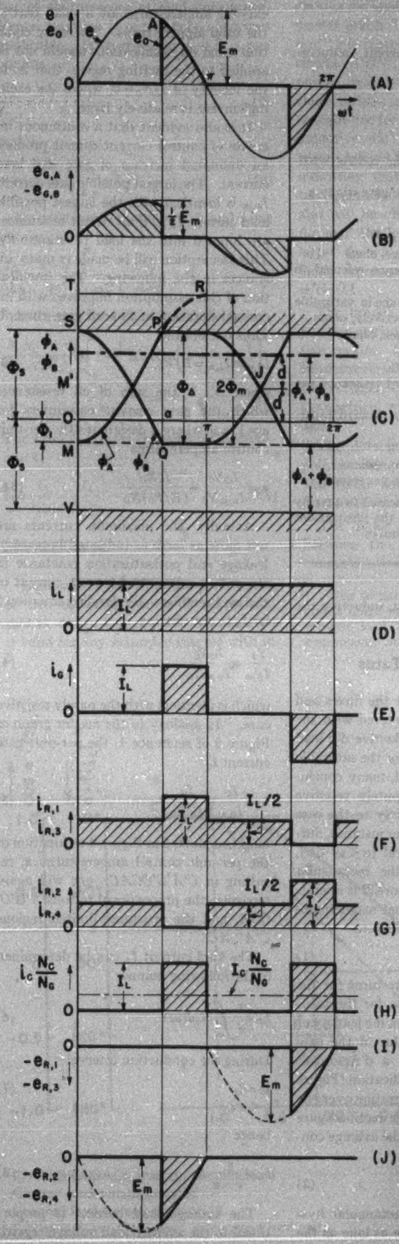


Figure 2 (left). Theoretical wave shapes. (A) Line voltage e, output voltage e_0. (B) Gate voltages. (C) Core fluxes. (D) Load current. (E) Gate current. (F) (G) Rectifier forward currents. (H) Control current. (I) (J) Rectifier reverse voltages

taking place. With two rectifier branches blocking, the impedance across *4* and *5* is now the forward resistance of two rectifier branches plus the impedance of the load. In view of the assumed long-time constant of the load, the load current will remain at essentially the same level. Gate current now can pass through the inductive load by an amount equal to the load current. This condition persists as long as the output voltage e_0 is larger than the voltage drop across the forward resistance of one rectifier branch. For zero rectifier forward resistance this condition terminates at $\omega t=\pi$, whereupon a new exciting interval commences.

It is during the conduction interval that the load is energized from the supply. During the exciting interval the load current is maintained at the expense of the magnetic energy stored in the load inductance L_L. The mechanical analogon is the free wheel, and for this reason one speaks of the free-wheeling action of the rectifier load circuit.

The currents $i_{R,1}$, and so forth, in the branches of the load rectifiers follow directly from the foregoing and their wave shapes are shown in Figures 2(F) and (G).

During the conduction interval one of the two cores of the saturable reactors is not saturated and will function as a current transformer. The result is a current flowing through the control circuit i_G which has the same wave shape as the gate current i_G, Figure 2(H).

For zero forward rectifier resistance and for negligible gate resistances, the load voltage e_L becomes identical with the output voltage e_0. During the conduction interval $\alpha<\omega t<\pi$, rectifiers REC_1 and

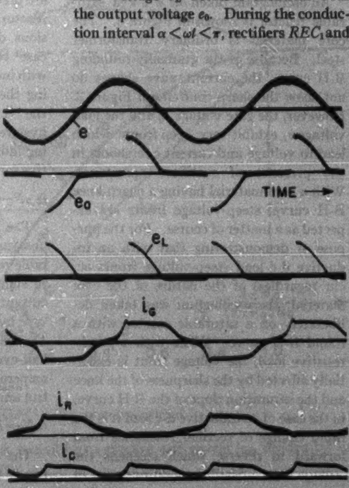


Figure 3 (right). Oscillogram. Line voltage e, output voltagee e_o, load voltage e_L, gate current I_g, rectifier forward current I_R, control current I_c

Nomenclature

$\mathfrak{a}$ = per-unit ampere-turns
E = supply voltage, volts, average
E_m = supply voltage, volts, maximum
e = supply voltage, volts, instantaneous
E_C = control voltage, volts, average
e_G, $e_{G,A}$, $e_{G,B}$ = gate voltage on saturable reactor A or B respectively
E_O = (nominal) output voltage, volts, average
e_O = (nominal) output voltage, volts, instantaneous
E_s = saturation voltage, volts, average, for one gate winding
$E_{s,1}$ = saturation voltage at frequency f_1, volts, average
f = supply frequency, cycles per second
$F(\mathfrak{a})$ = free-wheeling function, inverse of equation 12
I_C = control current, amperes, average
i_C = control current, amperes, instantaneous
I_G = gate current, amperes, average
$i_{G,m}$ = maximum gate current, amperes, average
$\mathfrak{i}_G$ = per-unit gate current
$i_{G,A}$, $i_{G,B}$ = gate current in saturable reactor A or B respectively, amperes, instantaneous
I_L = load current, amperes, average
$I_{L,e}$ = effective value of load current, amperes, rms
$I_{L,m}$ = maximum load current, amperes, average
$I_{L,\min}$ = minimum load current, amperes, average
$\mathfrak{i}_L$ = per-unit load current
K_{AT} = ampere-turn gain
K_E = voltage gain
K_I = current gain
K_P = power gain
N_C = number of turns of one control winding
N_G = number of turns of one gate winding
n = matching index
P_C = control power, watts
P_L = useful power in R_L, watts
R_C = control circuit resistance, ohms
R_C' = control circuit resistance referred to gate winding, ohms
R_G, $R_{G,A}$, $R_{G,B}$ = gate resistance in saturable reactor A or B respectively, ohms
R_G^* = effective gate resistance, ohms
R_L = load resistance, ohms
R_O = output resistance, ohms
R_R = normal rectifier forward resistance per branch, ohms
$\mathfrak{r}$ = per-unit normal rectifier forward resistance, ohms
t = time, seconds
α = starting angle, degrees or radians
ω = angular velocity, radians per second

Per-unit quantities are indicated in text by Old English letters, and in the illustrations by a bar underneath the symbol.

REC_3 carry the full load current, and the voltages $e_{R,2}$ and $e_{R,4}$ of the nonconducting rectifiers are equal to the negative load voltage $-e_L$, as an inspection of Figure 1 will disclose immediately. The rectifier voltages $e_{R,1}$, $e_{R,2}$, $e_{R,3}$, and $e_{R,4}$ are called reverse voltages and they are shown in Figures 2(I) and (J).

An oscillogram taken on a 2-kva saturable reactor is shown in Figure 3. The core material is ordinary transformer steel. Because of the gradually rounding B-H curve, the current wave shapes do not show the sharp contours of Figure 2. However, the gate voltage e_G and the load voltage e_L exhibit very steep fronts, which lead to voltage and current overshoots in the presence of rectifier capacitance. With a core material having a sharp-knee B-H curve, steep-voltage fronts are expected as a matter of course. For the purpose of demonstrating that with an inductive d-c load steep-voltage fronts occur regardless of the nature of the core material, the oscillogram was taken deliberately on a saturable reactor with a round B-H curve. While with a purely resistive load, the voltage front is essentially affected by the sharpness of the knee and the saturation slope of the B-H curve, in the case of an inductive d-c load it is the rapid change of rectifier resistance from forward to reverse which steepens the voltage front. Similar conditions exist in amplistat circuits with an inductive d-c load.

Law of Equal Ampere-Turns

In view of the fact that the fluxes and hence the voltages encountered with the saturable reactor with inductive d-c load are precisely the same as for the saturable reactor with resistive load, many conclusions derived from the purely resistive case[1] can be applied directly to the case with inductive d-c load. For instance, during the conduction interval ($\alpha<\omega t<\pi$), the vertical section of the rectangular hysteresis loop on which core B is operating indicates that a zero net magnetomotive force is required. Therefore

$$i_{G,B}N_G+i_CN_C=0 \tag{1}$$

The law of equal ampere-turns for the unsaturated core holds also for the saturable reactor with inductive d-c load. Let I_G signify the average value of the gate current, as measured by a d'Arsonval-type instrument after rectification (Figure 1); then it follows by integration over one half-cycle that the average rectified gate ampere-turns I_GN_G equal the average control ampere-turns I_CN_C

$$I_CN_C=I_GN_G \tag{2}$$

The assumption of a rectangular hysteresis loop is only tenable as long as the exciting ampere-turns are a small part of the total ampere-turns. Therefore equation 2 and its ramifications should not be applied to the exciting region, that is, to the vicinity of $I_CN_C=0$, where the exciting current is relatively large.

It is also evident that a continuous increase of control current cannot produce an unlimited increase of gate and load current. The largest possible gate current $I_{G,m}$ is identical with the largest possible load current $I_{L,m}$, if all circuit resistances are lumped into the load resistance R_L. This assumption will be made in many instances in the following. The justification of this assumption improves with increased duration of flow of gate current. Thus one obtains

$$I_{G,m}\equiv I_{L,m}=E/R_O \tag{3}$$

where R_O is the sum of all resistances which the gate current encounters for $\mathfrak{a}=1$, and where $\mathfrak{a}$ designates the per-unit control ampere-turns

$$\mathfrak{a}=\frac{I_CN_C}{I_{L,m}N_G}=\frac{I_CN_C}{(E/R_O)N_G} \tag{3A}$$

Actually, the maximum currents are not quite as large as indicated because of leakage and postsaturation reactance in the gate windings and reverse current of the load rectifier. By dividing equation 2 by $I_{G,m}$, one obtains

$$\frac{I_G}{I_{G,m}}=\frac{I_CN_C}{I_{G,m}N_G} \tag{4}$$

which is identical with the purely resistive case. In analogy to the master graph of Figure 5 of reference 1, the per-unit gate current $\mathfrak{i}_G$

$$\mathfrak{i}_G=\frac{I_G}{I_{G,m}}=\frac{I_G}{E/R_O} \tag{5}$$

has been plotted in Figure 4 as function of the per-unit control ampere-turns $\mathfrak{a}$, resulting in $C'A'G'OGAC$; one will again recognize the proportional regions $A'G'O$, OGA, and the resistance-limited regions $C'A'$, AC.

The load current I_L can be determined in the following manner

$$I_G=\frac{1}{\pi}\int_{\alpha}^{\pi} i_G d(\omega t) \tag{6}$$

During the conduction interval

$$\mathfrak{i}_G=I_L \qquad \alpha<\omega t<\pi \tag{7}$$

hence

$$I_L=\frac{\pi}{\pi-\alpha}I_G \tag{8}$$

The average load current is proportional to the average load voltage regardless of the presence of load inductance. Hence, equation 15 of reference 1 becomes applicable. With $\alpha=0$ for $I_L=I_{L,m}$, one obtains

$$\frac{I_L}{I_{L,m}}=\frac{1+\cos\alpha}{2} \tag{9}$$

and

$$\alpha=\arccos\left(\frac{2I_L}{I_{L,m}}-1\right) \tag{10}$$

By combining equations 8, 3, and 4 one obtains

$$\frac{I_G}{I_{G,m}}=\frac{\pi-\alpha}{\alpha}\,\frac{I_L}{I_{L,m}}=\frac{I_C N_C}{I_{L,m}N_G} \tag{11}$$

Substituting from equation 10 yields

$$\frac{I_C N_C}{I_{L,m}N_G}=\frac{I_L}{I_{L,m}}\left[1-\frac{1}{\pi}\arccos\left(2\,\frac{I_L}{I_{L,m}}-1\right)\right]$$
$$=\frac{2}{\pi}\,\frac{I_L}{I_{L,m}}\arcsin\sqrt{\frac{I_L}{I_{L,m}}} \tag{12}$$

The ratio $I_L/I_{L,m}$ is the per-unit load current $\mathfrak{i}_L$

$$\mathfrak{i}_L=\frac{I_L}{I_{L,m}}=\frac{I_L}{E/R_O} \tag{12A}$$

Master Graph

Equation 12 describes the performance of the saturable reactor in normalized, that is, dimensionless form, and therefore is valid for any saturable reactor with inductive d-c load. This graph is called the master graph. In graphical form, equation 12 is represented in Figure 4 by the curves *A'D'O, ODA*. In contrast to purely resistive load, the normalized characteristic (equation 12) is no longer a straight line for the saturable reactor with inductive d-c load. The slope of the curve changes from ∞ at 0 to zero at A, indicating that the theoretical ampere-turn gain also changes from ∞ to 0. It also will be observed that for inductive d-c load the load current is consistently larger than for resistive load, the latter being represented by the straight lines *A'O, OA*.

The branch *ODA* is again called the positive region, branch *OD'A'* the negative region. The negative region results from a reversal of the polarity of the control voltage. Curves *ODA* and *OD'A'* are symmetrical with respect to the ordinate axis.

The dots in Figure 4 show measured points, taken on saturable reactors with Mu-Metal cores, and a selenium-type load rectifier.

The major reasons for the deviations between the theoretical and practical characteristics are:

1. The actual B-H curve has a finite slope and a finite width. This results in an increased load current due to an exciting component, most prominent for $\mathfrak{a}=0$. It also calls for more than the proportional amount of d-c ampere-turns as $\mathfrak{a}$ increases.
2. Leakage and postsaturation reactances of the gate windings reduce the load current.
3. The forward resistance of the load rectifier is variable, decreasing with increasing load current.
4. The shunt path provided by the reverse resistance of the load rectifier causes a reduction of the load current.

Except for the vicinity of $\mathfrak{a}=0$, the experimental points, when based on a measured $I_{L,m}$ are approximated by the following empirical equation

$$\mathfrak{i}_L\approx 2\mathfrak{a}-\mathfrak{a}^2 \tag{13}$$

shown by crosses in Figure 4.

Returning to equation 12, it expresses $\mathfrak{a}$ as function of $\mathfrak{i}_L$. Usually, the inverse function is preferred, which on account of its transcendental character is only symbolically expressed

$$\mathfrak{i}_L=F(\mathfrak{a}) \tag{14}$$

where $F(\mathfrak{a})$ symbolizes such a function of $\mathfrak{a}$ that curve *ODA* results. Since the function $F(\mathfrak{a})$ reflects the free-wheeling action of the inductive load circuit, this function is called free-wheel function.

The control voltage follows simply from

$$E_C=I_C R_C \tag{14A}$$

The relation between starting angle α and per-unit load current $\mathfrak{i}_L$ in equations

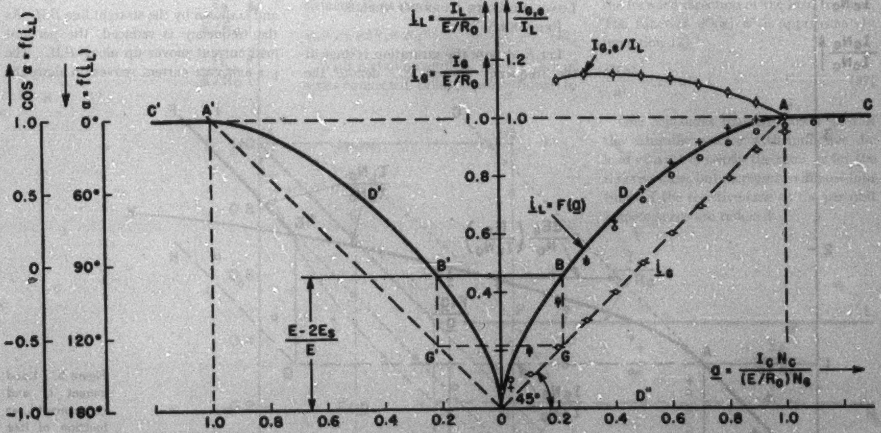


Figure 4. Master-graph

C'A'D'B'OBDAC: load current calculated
Circles: load current measured
Crosses: load current, equation 13
Circles with horizontal bars: gate current measured

O'A'G'OGAC: gate current calculated
α: starting angle
Circles with vertical bars: effective gate current $I_{G,e}$ divided by load current I_L

9 and 12(a) lead to the scales for cos α and α as shown in Figure 4. The notations f ($\mathbf{i}_L$) indicate that cos α and α are correlated to $\mathbf{i}_L$.

In the presence of overexcitation, namely, when the line voltage is so high that it would saturate the saturable reactors for $I_C=0$, the load current cannot fall below the minimum value $I_{L,\min}$ of equation 21 of reference 1. This is shown in the master graph of Figure 4 by the straight line BB' for the load current and GG' for the gate current. Points G and G' are obtained by projecting B and B' on the gate-current characteristic $A'G'OGA$. The regions BB' and GG' are called overexcited regions.

When dealing with temperature-limited reactors, the relation between the effective value of the gate current $I_{G,e}$ and the average value of the load current I_L is important. This relation has been established by measurement and is plotted in Figure 4 as function of the per-unit ampere-turns $\mathbf{a}$. If one wishes to determine $I_{G,e}$ for a given $\mathbf{i}_L$, one first obtains $\mathbf{a}$ from the measured curve represented by the plain circles and thence $I_{G,e}/I_L$.

Characteristics

As with the saturable reactor with resistive load, the characteristics of the saturable reactor with inductive d-c load can be derived from the master graph in Figure 4.

LOAD AND GATE CURRENT VERSUS CONTROL CURRENT

In view of the nonlinearity of the characteristic, the most practical way to obtain numerical values is to change the scales of the master graph. The ordinate $\mathbf{i}_L=1$ is changed to $I_L=E/R_O$, and all other ordinates are multiplied by the same ratio. The abscissa $\mathbf{a}=1$ is changed to $I_C=(E/R_O)$ (N_G/N_C) and all other abscissas are multiplied by the same ratio.

LOAD AND GATE CURRENT VERSUS SUPPLY VOLTAGE AND LOAD

Figure 5 illustrates the effect of the ratio E/R_O upon load and gate current. For a detailed discussion see Appendix I.

One concludes that the operation in the resistive-limited region OA is the same for resistive load and inductive d-c load. In the positive region, the load current of the saturable reactor with resistive load stays constant (AG), whereas the load current of the saturable reactor with inductive d-c load increases (AB). The rate of increase of the load current is smaller than the rate of increase of the ratio E/R_O. In the overexcited region BF, GH, the load currents for the resistive and inductive case are different, but their rate of rise with E/R_O is the same in both cases.

LOAD AND GATE CURRENT VERSUS FREQUENCY

Let E_s denote the saturation voltage at the frequency f, and $E_{s,1}$ denote the saturation voltage at the frequency f_1. Then

$$E_{s,1}/E_s=f_1/f \tag{23}$$

The frequency f_1 is now selected such that

$$2E_{s,1}=E \tag{24}$$

The frequency ratio f/f_1 for the boundary follows from equations 22, 23, and 24

$$f/f_1=1-F(\mathbf{a}) \tag{25}$$

For $f/f_1>1-F(\mathbf{a})$, the operation is in the positive region. The gate current follows from equation 4 and the load current from equation 15, indicating that the frequency has no effect on gate and load current. In Figure 6 the gate current is shown by GN which has the distance $\mathbf{a}$ from the abscissa axis. Intersection H with MK makes $JK=\mathbf{a}$. The vertical through JH intersects $F(\mathbf{a})$ at P, hence the horizontal PL represents the per-unit load current. The distance $PQ=1-F(\mathbf{a})$: because of the geometry, $PQ=OR=f/f_1$. In view of equation 25, this value of f/f_1 is the boundary point, and hence a vertical through RG, intersecting the extension of PL at B, constitutes the boundary point B for the load current. Point B also is situated on MK. The corresponding boundary point for the gate current is G.

During overexcitation, equation 20 holds, resulting in

$$\mathbf{i}_L=1-\frac{f}{f_1} \tag{26}$$

and is shown by the straight line BM. As the frequency is reduced, the per-unit load current moves up along BM. The per-unit gate current moves up along the

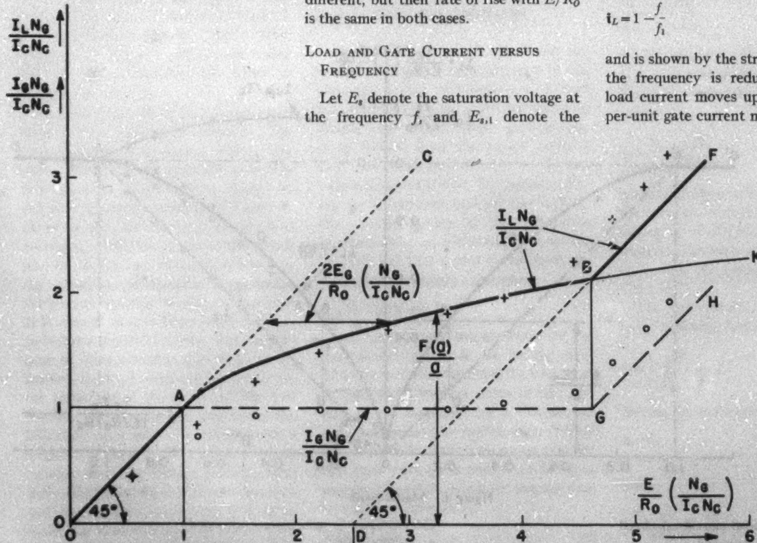


Figure 5. Load current I_L and gate current I_G as function of line voltage E and output (load) resistance R_O. Crosses and circles are measured points

curve *GM*, which is symmetrical to $F(\mathbf{a})$ with *KM* as the axis of symmetry.

Measured points are shown in Figure 6 by circles and crosses. Similar to Figure 5, the best accuracy is associated with the gate current in the proportional region. The test points were taken on a saturable reactor with Mu-metal core, $N_G=592$, $N_C=600$, $I_C=19$ milliamperes, $R_o=402$ ohms, $2E_{s,1}=E=36$ volts and $f_1=52.2$ cycles per second.

Gain

When dealing with the amplifying property of saturable reactors, one of the most important factors is the degree of amplification, which is called the gain. There are several types of gain to be considered.

AMPERE-TURN GAIN

By definition and in view of equations 14 and 3(A), one obtains

$$K_{AT}=\frac{d(I_LN_G)}{d(I_CN_C)}=\frac{d\mathbf{i}_L}{d\mathbf{a}}=F'(\mathbf{a}) \tag{27}$$

In contrast to the saturable reactor with resistive load, the ampere-turn gain varies as a function of the per-unit control ampere-turns. The theoretical variation (equation 12) is between ∞ and 0. Practically, however, the variation is not as drastic. One obtains from equations 27 and 13

$$K_{AT}=F'(\mathbf{a})\approx 2-2\mathbf{a} \tag{28}$$

One realizes that for $\mathbf{a}=0.5$ the ampere-turn gains of the saturable reactor with resistive and inductive d-c load are the same, namely $K_{AT}=1$. For $\mathbf{a}<0.5$ the ampere-turn gain of the inductive d-c case is larger, and for $\mathbf{a}>0.5$ the gain is smaller than for the resistive case. The size of the reactor, its core material, and the number of turns are, in a first approximation, immaterial to the ampere-turn gain.

POWER GAIN

The power gain is defined by

$$K_P=P_L/P_C \tag{29}$$

By applying equations 3(A) and 12(A), one obtains

$$\frac{P_L}{P_C}=\frac{\mathbf{i}_L{}^2R_L}{\mathbf{a}^2R_C'}=\frac{F^2(\mathbf{a})}{\mathbf{a}^2}\,\frac{R_L}{R_C'} \tag{30}$$

where

$$R_C'=R_C(N_G/N_C)^2 \tag{31}$$

Let the effective gate resistance $R_G{}^*$ be defined as the sum of the gate resistances for the series-connected saturable reactor, and one-half of one gate resistance for the parallel-connected saturable reactor. With R_R denoting the forward resistance of one branch of the load rectifier, one obtains for the total circuit resistance R_O at $\mathbf{a}=1$

$$R_O\approx R_L+2R_R+R_G{}^* \tag{32}$$

With metallic rectifiers, the number of series-connected cells is proportional to the load voltage, and the number or area of parallel-connected cells is proportional to the load current, hence the forward resistance of the load rectifier for $\mathbf{a}=1$ is essentially proportional to the load resistance

$$2R_R=\mathbf{r}R_L \tag{33}$$

$\mathbf{r}$ is called the per-unit rectifier forward resistance. In case of new selenium rectifiers, $\mathbf{r}$ is about 0.15 at 20 degrees centigrade. The resistance of the inductive load, therefore, appears enlarged by $\mathbf{r}R_L$ as compared with the resistive case.

Defining the load-matching index n by

$$n=(1+\mathbf{r})R_L/R_G{}^* \tag{34}$$

one obtains the power gain from equation 30

$$K_P=\frac{F^2(\mathbf{a})}{\mathbf{a}^2}\,\frac{R_G{}^*}{R_C'}\,\frac{n}{1+\mathbf{r}} \tag{35}$$

It can be shown that, for the same temperature rise in gate and control winding, the power reaches an optimum for $R_G{}^*=R_C'$ (including winding resistance only), and therefore

$$K_P=\frac{F^2(\mathbf{a})}{\mathbf{a}^2}\,\frac{n}{1+\mathbf{r}} \tag{36}$$

For equal temperature rise in gate and control winding, the power gain is proportional to the matching index, as in the case with resistive load. The gain is reduced by $1/(1+\mathbf{r})$ because of the loss in the forward direction of the load rectifier. The function $F^2(\mathbf{a})/\mathbf{a}^2$ is approximately (equation 13)

$$\frac{F^2(\mathbf{a})}{\mathbf{a}^2}\approx(2-\mathbf{a})^2 \tag{37}$$

One concludes that the power gain of the saturable reactor with inductive d-c load at $\mathbf{a}=1$ is nearly the same as for the resistive case, but increases to almost four times of the resistive case as the per-unit ampere-turns are reduced.

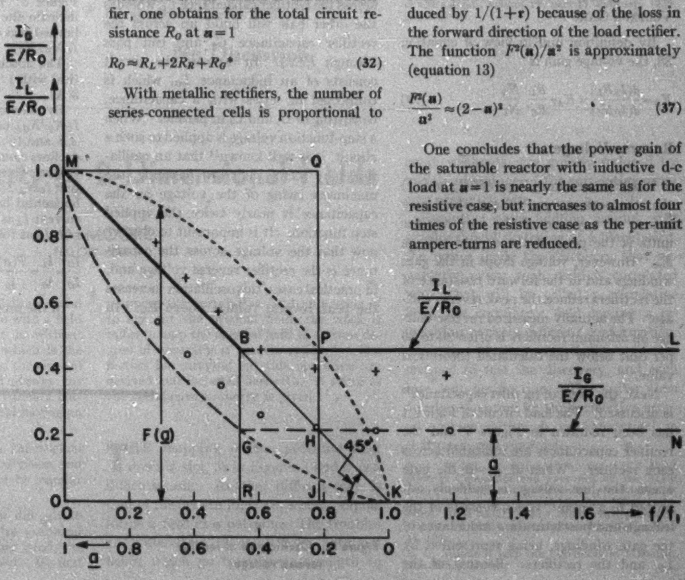


Figure 6. Load current I_L and gate current I_G as function of the line frequency f. Crosses and circles are measured points

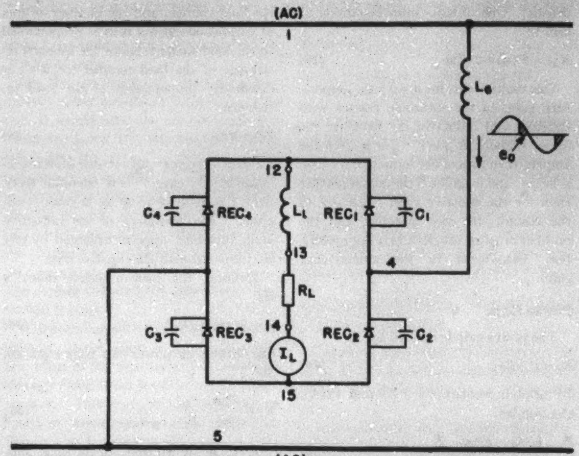


Figure 7. Oscillatory circuit, responsive to shock excitation by the steep front of output voltage e_0

CURRENT GAIN

By definition and in view of equations 14, 3(A), and 12(A) one obtains

$$K_I = \frac{dI_L}{dI_C} = K_{AT}\frac{N_C}{N_G} \tag{38}$$

VOLTAGE GAIN

By definition and in view of equation 38, the voltage gain is

$$K_E = \frac{d(I_L R_L)}{d(I_C N_C)} = K_{AT}\frac{R_L}{R_C'}\,\frac{N_G}{N_L} \tag{39}$$

Reverse Voltage

It will be seen from Figures 2(I) and (J) that the peak reverse voltage which may appear at any one of the four rectifier units is the peak of the supply voltage E_m. However, voltage drops in the gate windings and in the forward resistance of the rectifiers reduce the peak reverse voltage. The actually measured reverse voltage on selenium rectifiers is often 20 to 30 per cent below the calculated theoretical value.

Next, the effect of rectifier capacitance[10] is discussed. The load circuit of Figure 1 has been redrawn in Figure 7 and the rectifier capacitances are indicated across each rectifier. When at $\omega t = \alpha$ the gate opens, the line voltage is suddenly admitted to a circuit, which consists of the leakage and postsaturation inductance of the gate windings, being represented by L_G, and the rectifiers. Because of the large inductance L_L of the load, the latter will not take part effectively in the transient described in the following. The positive surge coming down through L_G at $\omega t = \alpha$ will divide at point *4*. One-half of it will pass through REC_1 and appear across the rectifier capacitance C_4, whereas the other half will build up across the rectifier capacitance C_2 and can pass through REC_3. In essence, the circuit consists of an inductance L_G, which is connected in series with a capacitance, consisting of C_2 and C_4 in parallel. When a step-function voltage is applied to such a circuit, it is well known[11] that an oscillation may occur. For low damping, the maximum swing of the voltage on the capacitance is nearly twice the applied step function. It is important to observe now that the voltage across the capacitance is the rectifier reverse voltage and, in practical cases, this oscillation increases the peak reverse voltage above E_m. In a typical practical case with selenium rectifiers, the peak reverse voltage is about 1.5 E_m. The reverse voltage as a function of time is shown in the cathode-ray oscillogram of Figure 8. The frequency of oscillation depends, among other things, on the per-unit ampere turns, and is usually in the kilocycle range. As a further result of the oscillatory nature of the circuit, oscillations similar to the reverse voltage also occur in the currents of the saturable reactor.

To reduce the peak reverse voltage, a resistance can be connected in shunt with the inductive load. A resistance of approximately $5R_L$ is usually sufficient to reduce the overshoot of the reverse voltage to about 5 per cent. Another means for eliminating the reverse voltage overshoot consists in a capacitor connected across the d-c terminals of the load rectifier.

Appendix I

The ratio of line voltage E over output resistance R_O is raised from zero, while the control current I_C is held constant; it is assumed again that all circuit resistances can be lumped with the load resistance. For small values of the ratio $E/R_O (E/R_O < I_C N_C/N_G)$, the operation is in the resistance-limited region AC of Figure 4. Per-unit currents $\mathbf{i}_L$ and $\mathbf{i}_G$ remain constant at unity, and hence I_L and I_G rise linearly with the ratio E/R_O. This is shown by OA in Figure 5; points A in both illustrations indicate the boundary between resistance-limited region and positive region $(E/R_O = I_C N_N/N_G)$.

The abscissa $(E/R_O)(N_G/I_C N_C)$ is identical with $1/\mathbf{u}$ as can be seen from equation 3(A).

By increasing the ratio $E/R_O (E/R > I_C N_C/N_G)$, one traverses the positive regions AB and AG. The per-unit gate current $\mathbf{i}_G$ declines along the dashed straight line AGO (Figure 4), resulting in a constant gate current $I_G N_G/I_C N_C = 1$; this is illustrated by the horizontal line AG in Figure 5. The load current I_L is determined from the gate current I_G as follows

$$\frac{I_C}{I_G} = \frac{\mathbf{i}_L}{\mathbf{i}_G} = \frac{F(\mathbf{u})}{\mathbf{i}_G} \tag{15}$$

In the proportional region

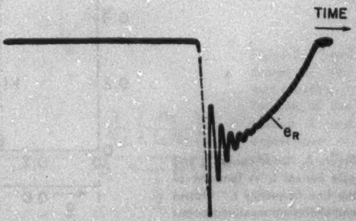


Figure 8. Oscillogram of rectifier reverse voltage

$$i_G=\boldsymbol{a} \tag{16}$$

hence

$$I_L=I_G\frac{F(\boldsymbol{a})}{\boldsymbol{a}} \tag{17}$$

The function $F(\boldsymbol{a})/\boldsymbol{a}$ can be determined numerically from equation 12 or graphically from Figure 4, and it is plotted in Figure 5 as curve *ABK*. For small values of $\boldsymbol{a}$, the function $F(\boldsymbol{a})$ is approximated by

$$\frac{F(\boldsymbol{a})}{\boldsymbol{a}}\approx\left(\frac{\pi}{2}\right)^{1/3}\boldsymbol{a}^{-1/3} \tag{18}$$

Since $I_GN_G/I_CN_C=1$ in the proportional region, the load current becomes

$$\frac{I_LN_G}{I_CN_C}=\frac{F(\boldsymbol{a})}{\boldsymbol{a}} \tag{19}$$

In the absence of the saturable reactors, the load current would increase linearly from *A* to *C* on an extension of *OA*. With a resistive load, the load current remains constant *AG*. With inductive d-c load, the current responds to changes of the ratio E/R_O along curve *AB*, which is an intermediate between *AC* and *AG*.

One concludes that the load current does not remain constant as in the case of the saturable reactor with resistive load, but increases with the ratio E/R_O, although at less than proportionality.

The horizontal distance between *AC* and the ordinate axis corresponds to the output voltage E_O; hence the horizontal distance between *AB* and *AC* corresponds to the sum of the gate voltages. By definition, the sum of the gate voltages cannot become larger than the saturation voltages $2E_s$; therefore the intersection at *B* between *DF* with *AK* terminates the positive region, corresponding to point *B* of the master graph in Figure 4. The gate current reaches the end of the proportional region at a ratio E/R_O, as represented by point *G* in Figures 3 and 5. The ratio E/R_O, corresponding to the boundary point *B* between positive and overexcited region, follows from equation 21 of reference 1 and equations 15 and 16.

$$I_L=\frac{E-2E_s}{R_O} \tag{20}$$

$$I_L=\frac{I_CN_C}{N_G}\frac{F(\boldsymbol{a})}{\boldsymbol{a}} \tag{21}$$

$$\frac{E}{R_O}\frac{N_G}{I_CN_C}=\frac{F(\boldsymbol{a})}{\boldsymbol{a}}+\frac{2E_s}{R_O}\frac{N_G}{I_CN_C} \tag{22}$$

This proves again that the intersection *B* of *DF* with *AK* marks the boundary.

If the ratio E/R_O is raised above the boundary value of equation 22, the operation occurs in the overexcited region. The load current obeys equation 20. Both the gate current and the load current increase at the same rate (*BF* and *GH* shown in Figure 5).

The circles and crosses in Figure 5 indicate measured points. The test was taken on a saturable reactor with Mu-metal core; the parameters were: $N_G=592$, $N_C=600$, $I_C=40$ milliamperes, $R_O=402$ ohms, $2E_s=41.3$ volts. With the performance of the saturable reactor established as function of the ratio E/R_O, one can immediately obtain the performance as function of either *E* or R_O by freezing the other. Numerical values for I_L and I_G are obtained by multiplying the ordinates by I_CN_C/N_G. For a fixed R_O, numerical values for *E* are obtained by multiplying the abscissas by $R_OI_CN_C/N_G$, and for a fixed *E* numerical values for $1/R_O$ are obtained by multiplying the abscissas by I_CN_C/N_GE.

References

1. **Series-Connected Saturable Reactor with Control Source of Comparatively Low Impedance, H. F. Storm.** *AIEE Transactions*, volume 69, part II, 1950, pages 756–65.

2. **A New Theory of the Magnetic Amplifier, A. G. Milnes.** *Proceedings*, Institution of Electrical Engineers (London, England), volume 97, part II, August 1950, pages 460–83.

3. **General Characteristics of Magnetic Amplifiers, L. A. Finzi, D. C. Beaumariage.** *AIEE Transactions*, volume 69, part II, 1950, pages 919–27.

4. **Universal Curves for D-C Controllable Reactors, Walter C. Johnson, B. C. Merrell, R. E. Alley, Jr.** *AIEE Transactions*, volume 68, part I, 1949, pages 31–40.

5. **Theory of D-C Excited Iron-Core Reactors and Regulators, A. Boyajian.** *AIEE Transactions*, volume XLIII, June 1924, pages 919–36.

6. **Bibliography of Magnetic Amplifier Devices and the Saturable Reactor Art, James G. Miles.** *AIEE Transactions*, volume 70, part II, 1951, pages 2104–23.

7. **Wave Forms (book), B. Chance.** McGraw-Hill Book Company, New York, N. Y., 1949, page 364.

8. **Transient Response of Saturable Reactors with Resistive Load, H. F. Storm** *AIEE Transactions*, volume 70, part I, 1951, pages 95–102.

9. **Series-Connected Saturable Reactor with Control Source of Comparatively High Impedance, H. F. Storm.** *AIEE Transactions*, volume 69, part II, 1950, pages 1299–1309.

10. **Metal Rectifiers (book), H. K. Henisch.** Clarendon Press, Oxford, England, 1949, page 35.

11. **The Radio Engineers Handbook (book), K. Henney.** McGraw-Hill Book Company, New York, N. Y., third edition, 1941, page 128.

12. **The Behaviour of a Series-Transductor Magnetic Amplifier with Directly-Connected or Rectified-Fed Loads, A. G. Milnes.** *Proceedings*, Institution of Electrical Engineers (London, England), volume 99, part II, February 1952, pages 13–23.

No Discussion

A New Property of 2-Dimensional Fields

A. D. MOORE
FELLOW AIEE

Synopsis: A new property of 2-dimensional fields is presented. It has to do with a field in a nearly closed area that is without sources, a typical example of which is the flux in a slot. This will be called the original field. Any such original field always has correlated with it what will be called a correlative field, the correlative field having an entirely different conformation.

The correlation appears when the original field is mapped by partial subdivision and the correlative field is mapped by regular subdivision.

Such original fields are very difficult to map when attacked directly by graphical mapping methods. On the other hand, the correlative field is easily mapped. By first mapping the correlative field and using it to lay out the original field, the work it takes to map the original field to a given degree of accuracy is greatly reduced. The factors for carrying out this operation are derived and included herewith. A proof of the correlative property is given.

THE mapping of the potential field down a slot, as in Figures 1 and 2, is a difficult task. Another difficult map is that of the field in Figure 6, where a dipole forms a field in a polygon. The troubles encountered in mapping such potential fields, which go from great strength to great weakness, are greatly reduced by taking advantage of the correlative property presented here.

Recently, while fiddling around with graphical field sketching at a blackboard, the writer made the discovery that apparently every such field is correlated with what will be called a correlative field, the correlative field having a related but entirely different conformation. A number of cases were then graphically mapped to test the discovery, and each case lent verification. The writer then developed a proof for the correlative property which, although not analytical in character, was rigorous.

Paper 52-240, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Summer General Meeting, Minneapolis, Minn., June 23–27, 1952. Manuscript submitted January 30, 1952; made available for printing May 14, 1952.

A. D. Moore is with the University of Michigan, Ann Arbor, Mich.

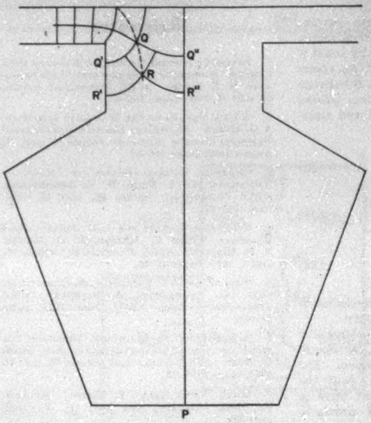


Figure 1 (left). Map started in the high-strength region

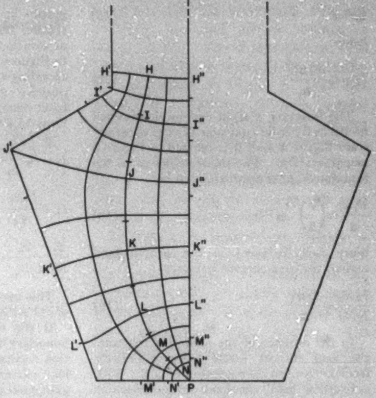


Figure 3 (right). The correlative field, mapped by regular subdivisions

Since it seemed somewhat improbable that a new property of 2-dimensional potential fields should turn up at this late date, a preliminary draft of the present paper was given private distribution to a number of workers, some of whom are familiar with field theory and others with mapping methods. All who replied found the correlative property to be new to them.

Also, two of these men, Dr. William R. Smythe of California Institute of Technology and Mr. G. L. Godwin of Westinghouse Electric Corporation, went into it far enough to work out proof by means of the Schwarz-Christoffel transformation. It is Dr. Smythe's analytical proof that is included in Appendix I.

The correlative property appears when the original field (such as slot flux) is mapped by partial subdivision, and the correlative field is mapped by regular subdivision. The property is not only interesting in its own right; additionally, by taking advantage of it, the original field may be mapped far more easily.

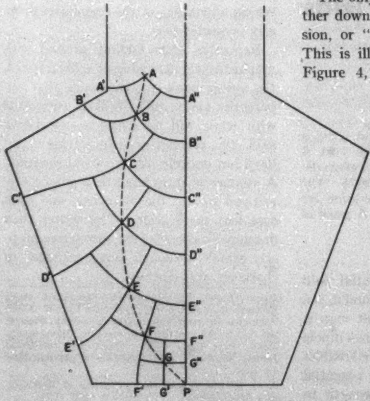


Figure 2 (left). The original field, mapped by partial subdivisions

The Original Field

A slot of arbitrary shape has been drawn in Figure 1. By graphical methods, a regular array of curvilinear squares has been drawn in the gap and over the slot mouth. These are high-strength regions. But as we attempt to carry the map down the slot into the weak-flux region, we have our troubles. In this incomplete map, the square $PQ'QQ''$ has been regularly subdivided once into four squares.

The only feasible routine for going further down is to resort to partial subdivision, or "subdivision toward a corner." This is illustrated in a rectilinear field, Figure 4, right half. Let us identify squares by giving only the two letters on opposite corners. The large square AP has been regularly subdivided once into four squares. Next, instead of subdividing all four squares again, only BP has been subdivided; next, only CP is subdivided, and so on. There is ordinarily no advantage gained by so treating a rectilinear field. But this routine is an imperative in handling many weak-field problems.

When applied to the slot as in Figure 2, we see partial subdivision carried out; the successive subdivisions are always toward the corner P. By virtue of the method of subdivision and of the properties of curvilinear squares, certain important advantages are achieved. For one thing, the flux quantities arriving on $A'B'$, $B'C'$, $C'D'$, and so on, are as $1:(1/2):(1/4)\ldots$ The potential differences from A'' to B'', B'' to C'', C'' to D'', and so forth, vary likewise.

This map, being finished, is easy to look at; but when directly attempted by graphical methods, it is a difficult problem indeed.

Now take note of the dashed bisector

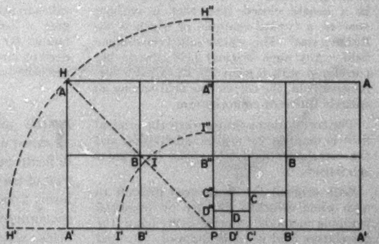


Figure 4 (right). Special case, used for determination of factors. Original field rectangular, solid lines. Correlative field radial-circular, dash lines

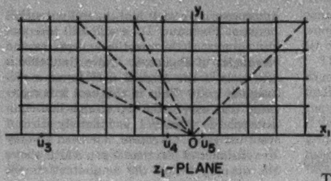
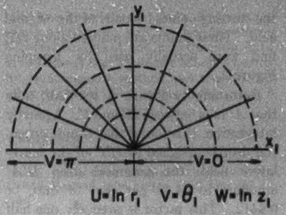


Figure 5 (left). See Appendix I

Figure 7 (right). See Appendix I



line *AP*. This curved line bisects all of the angles through which it passes, these angles all being, of course, 90-degree angles in the orthogonal system. If this bisector line could be laid out first, the partial subdivision mapping job would be greatly facilitated. As will be seen, the bisector can be had, in advance.

The Correlative Field

In Figure 3 the slot outline has been copied except that we now extend the slot mouth upward towards infinity. The center line, which was a flux line in the original field, is retained as a flux line in the correlative field. But the slot boundary *A'P* of Figure 2, an equipotential in the original field, is now made into a flux line *H'P* of Figure 3 in the correlative field. The kind of field we now get, for example, could be set up starting with a thin, uniform conducting sheet; cutting out of it a piece the shape of the slot with the neck extended upward for a considerable distance; attaching one electrode across the top of the neck; and attaching another, and small electrode, at point *P*.

This new field, which will turn out to be the correlate of the original field though very different from it, may be mapped easily in regular fashion as shown. When so mapped, *HP* automatically is the midline; in the left half of the slot it divides the current or flux into equal parts.

Relationships

A first proposition is that the midline of the correlative field is identical with the bisector line of the original field.

If this correlation can hold, then subsidiary correlations can be expected, some of which might prove to be useful. Such is the case. Certain factors can be found by which the correlative field can be made to help lay out the original field. These factors are determined most easily by using a field of simple conformation, such as a rectilinear field.

In Figure 4, a rectilinear field (solid lines) is taken as the original field. The *AP* square at the left has been subdivided once. The lettering here is the same as that used in the original field of Figure 2. The correlative field (dash lines) will consist of radial and circular lines. Let the correlative field be started with the arc *H'HH"* through point *A*, *A* and *H* being designations for a point common to both fields. It is a matter of simple integration (see Appendix I) to find the radius *IP*, such that squares *H'I* and *H"I* are constructed as true curvilinear squares. Note that *B* falls outside of *I*. Again, by integration it is found that going from *H* to *B* takes in 88.25 per cent of the correlative field's potential drop from *H* to *I*.

A second proposition is that this factor holds for any original field and its correlative field. Applied to Figure 3, let it be decided to start a pair at *H*. That is, the *A'A* and *A"A* lines of Figure 2, forming an original field pair, are going to be located, in part, by drawing the pair's intersection *A* at point *H* of Figure 3. Then go down along *HI* to a point where we have covered 88.25 per cent of the potential drop along *HI*. This is not done linearly, of course, but by appropriate interpolation within the map. Subdivision of square *H'I* furnishes a scale for such an interpolative step. The point so found locates *B* of Figure 2, or the point at which to begin the next pair of the original field. If we assign to the squares *H'I*, *I'J*, and so forth, unit potential drop each, then we need to step off from *H*, potential drops each of 0.8825 value along *HP*, to locate

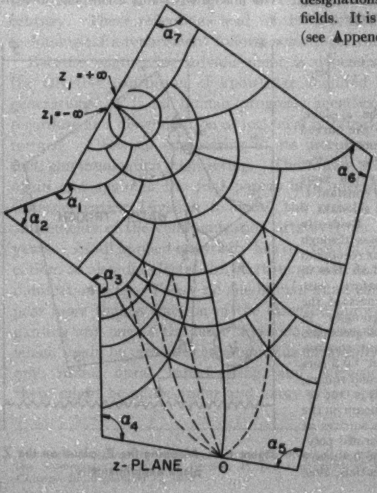


Figure 6 (left). See Appendix I

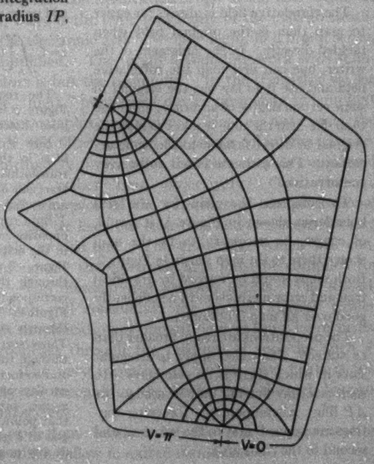


Figure 8 (right). See Appendix I

the starting points for pairs of the original field. These points are marked along HI, find were actually used for mapping Figure 2.

Returning to Figure 4, left half, note that A' falls between H' and I'. Again by integration, it is found that if unit potential drop is assigned to $H'I'$ of the correlative field, the potential drop $H'A'$ amounts to 44.12 per cent of that, or 0.4412. This factor is precisely one-half of the other factor found above.

A third proposition is that this second factor also holds for any field and its correlative field. Applied to Figure 3, we go along $H'P$ and starting at H', step off 0.4412; the next step has the value 0.8825 or the total counted from H' is 0.4412 plus 0.8825; the next point to mark, as a total from H', is 0.4412 plus 2(0.8825); and so on. These points are marked. The same process applies to stepping off along $H''P$. Thus, the landing points for the feet of the pairs of the original field, Figure 2, are determined.

With the starts of pairs and foot locations thus found, the original field can be sketched in with facility and considerable accuracy.

Still other subsidiary relationships can, of course, be found, some of which might be useful. For example, if it were now desired to subdivide once each of the squares of Figure 2, additional factors can be found by which to locate the starts for the new pairs along AP and for locating the landing points for these pairs along $A'P$ and $A''P$.

Mapping

The correlative field is always far easier to map than is the original field when tackled directly. In the judgment of the writer, one can first map the correlative field and use it to lay out the original to some certain degree of accuracy, all in less than one-fifth the time it takes to map the original in direct fashion to the same accuracy. This estimate is felt to be very conservative.

To complete the map for the real boundaries shown in Figure 1, it would, of course, be necessary to fit the map shown there to the map shown in Figure 2. Both maps would be warped by graphical trial and error methods until a sufficiently good over-all map is obtained.

The propositions are really broader than as stated. They also apply to irregular slots in which there is no symmetry. In such case, any flux line may be used as the AP line; whereupon, the foregoing routines are applied, first to the left area, and second to the right area.

Appendix I

INTEGRATIONS (FIGURE 4)

To make $H'I$ a curvilinear square

$$\int^{R_2}\frac{dr}{(\pi/4)r}=\frac{4}{\pi}\log_e\frac{R_2}{R_1}=\frac{4}{\pi}\log_e\frac{HP}{IP}=1,$$

$$\frac{HP}{IP}=2.1933$$

Take potential drop $HI=1$. Then

$$\text{Potential drop } HB=\frac{4}{\pi}\log_e\frac{HP}{BP}=\ldots$$

$$=\frac{4}{\pi}\log_e 2=0.8825$$

$$\text{Potential drop } H'A'=\frac{4}{\pi}\log_e\frac{H'P}{A'P}$$

$$=\frac{4}{\pi}\log_e\frac{H'P}{\left(\frac{H'P}{\sqrt{2}}\right)}=\frac{4}{\pi}\log_e\sqrt{2}=0.4412$$

PROOF OF THE CORRELATIVE PROPERTY

The writer is very grateful to Dr. William R. Smythe for developing the following analytical proof, and sending it to the writer by letter. It is included here with his permission.

The correlative property has a sound theoretical basis which is evident by examining the Schwarz-Christoffel transformation. This transforms the upper half of the z_1-plane, bounded by the real axis and the infinite arc into the interior of a polygon in the z-plane. If the polygon is of finite size, it will be bounded entirely by the deformed real axis $y_1=0$ of the z_1-plane, and there will be a 2-dimensional dipole singularity where the ends of this axis join. If the polygon is not closed, portions of its boundary at infinity may be formed by expanding or contracting the original infinite arc of the z_1-plane.[1] Consider a closed polygon. To get it we apply the following transformation in which the interior angles add up to $n\pi$.

$$z=C\int\left[(z_1-u_1)^{\frac{\alpha_1}{\pi}-1}(z_1-u_2)^{\frac{\alpha_2}{\pi}-1}\ldots\right]dz_1$$

The rectangular grid of the z_1-plane, Figure 5 is now distorted into the curved lines shown in the z-plane, Figure 6.

Now if we plot the field $W=U+jV=\ln z_1$ in the z_1-plane and apply the same transformation, then the lines of constant U start out as circles and those of constant V are radial lines, see Figure 7. Evidently V-lines can be chosen which pass through opposite corners of sets of similar rectangles in the original grid, shown as dash lines in Figure 5. They must, therefore, pass through the corresponding corners of the corresponding curvilinear rectangles in Figure 6. Those starting at 45 degrees pass through the corners of adjacent squares. Those starting at $\tan^{-1} 2$ or $\tan^{-1} 0.5$ pass through the corners of every second square, and so forth. Observe that this is true regardless of where the point is chosen on the x_1-axis, provided one starts his squares at that point. Thus in the transformed polygon of Figure 8 we can split the boundary at any point into different potentials, even at a corner if we wish, and the equipotentials between $V=0$ and $V=\pi$ will still intersect the corners of the curvilinear squares and rectangles of Figure 6. Note that this is true only if the original field forms a rectangular grid. Thus in this case it is rigorously true only if the top straightline equipotential in Figure 1 is infinitely distant from the gap. Of course, one need go only to a distance of less than a gap width above the opening before the equipotentials are essentially straight.

It is well known that any infinitesimal area dS_1 in the z_1-plane transforms into an area dS in the z-plane without change of shape but with a magnification dz/dz_1 and a rotation. The ratio dz/dz_1 is a function of position only so the relative position and size of the superimposed infinitesimal squares of the original and correlative fields are identical at the corresponding points in the z- and z_1-planes. This justifies the curvilinear square interpolation used in the second proposition. In regions where the transformation strongly distorts the original field, the position ratio (0.8825 in proposition two) will be the same in the original and transformed fields only when the squares are taken very small.

Reference

1. STATIC AND DYNAMIC ELECTRICITY (book), William R. Smythe. McGraw-Hill Book Company, Inc., New York, N. Y., 1950, pages 82–99.

Discussion

John F. H. Douglas (Marquette University, Milwaukee, Wis.): The method of mapping and the proof submitted in this paper appear to be valuable extensions of Maxwell's theorem II[1] which describes the general process of making a new and different

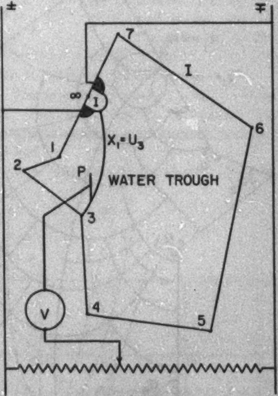


Figure 1. Mapping the Z_1 plane on the Z plane experimentally

field map from an old one. To construct a map like that shown in Figure 6 of the paper, a very short and convenient experimental method is one that uses an electrode shaped like a clover leaf. The clover leaf electrode, the water trough, the electronic voltmeter, and the potentiometer are shown in Figure 1 of the discussion. Wood or insulation is marked with the letter I, the letter P is a probe. The voltmeter is denoted by V. The potentials at the corners 1–7 give the functions U_1 to U_7 appearing in the last equation of the paper. By interchanging wood and metal on the boundary, the other set of lines can be mapped.

The great difficulty in solving the last equation of the paper analytically is in finding the roots U_1 to U_7 inclusive. Once these are found, the singular point marked O in Figure 6 may be found by a simple graphical integration. The resistance of the field for any mode of electrode location then becomes a simple analytical problem. It may be said that this method has been applied to a boundary for which the analytical solution is known with a close agreement between analytical and experimental data. It is believed that Professor Moore's beautiful permanganate method also could be applied to obtain the patterns in Figures 3, 6, and 8.

REFERENCE

1. ELECTRICITY AND MAGNETISM, volume 1 (book), James Clerk Maxwell. Clarendon Press, Oxford, England, third edition, 1891, page 287.

A Milestone in the Communication Arts

F. B. BRAMHALL
FELLOW AIEE

THE Centennial which engineers now celebrate marks not the end of a century of engineering but rather the 100th anniversary of the organization of engineers for the advancement of applied science. Truly engineering began at man's emergence from the very primitive state. No doubt the first use of fire was looked upon with as much awe and apprehension as is the control of nuclear fission today. To fix in mind the rate of scientific progress, we need only recall that until Lord Rayleigh it was quite conceivable that one mind might encompass and master all of science. No one realizes better than the engineer how futile it would be for one man to attempt to master all of science today. No one realizes better than the engineer the high degree of specialization that has developed in the mechanical arts during the first half of the present century. These reflections lead to the speculation that technological advancement follows some logarithmic law.

Because engineering advancement is dependent upon the effective utilization of knowledge acquired by past generations, as is all human progress, agencies for the recording and dissemination of technical information were formed. Electrical engineering at its inception was, in fact, communication engineering. This fact is attested to again and again in the four papers here offered by the Communication Division to detail the growth and advancements of the communication arts over the last 100 years. As all applied electricity has its roots in communication, so also did electronics start its growth in the communication arts. Today no one human mind can encompass more than a fraction of the essential knowledge regarding any one of the basic divisions of communication which might be presently catalogued as telegraphy, telephony, radio, aural broadcasting, and television. To these might be added radar and many other applications of electricity and electronics which have as their primary purpose the transportation of information. We may properly expect expansion of the electronic arts to continue at a rapidly accelerating pace.

The inspiration of that great host of engineers who devote their lives to communication is the realization that their contributions pave the way for progress of the human race. Communication must precede all business and social intercourse. One day we foresee communication providing such free and extensive exchange of information as will bring understanding among all nations and races, a common language, and universal peace.

The occasion to pause, reflect, and modify the course as seems indicated is valuable and welcome. This Centennial offers such an opportunity. While greater and greater complexity in communication may seem the trend at present, there is the hope that many of the instrumentalities of the electronic arts now may begin, in a broad sense, to be simplified. The recent astounding work with semiconductors is but the beginning. In the transistor we see a simple, "inert," permanent device promising to supplant a far more complex mechanism. May we not expect the oscillator, the filter, or even the whole color television receiver ultimately to so evolve that they may be considered subjectively rather than objectively? Only through continued expansion of the activities of our professional societies may we keep our scientists and engineers working effectively to this end. The knowledge which will continue to expand the electronic arts must be compiled for reference and disseminated for easy access. The knowledge which will consolidate the ground already gained through simplification of present devices must be recorded and broadcast through the medium of societies such as AIEE. The very incentive to invade new fields is fostered by the recognition the worker gets through his professional society.

The milestone marked by the four communication papers which follow is indeed significant.

Essential substance of the remarks made by Mr. Bramhall in introducing the the Communication Division Session at the Centennial of Engineering, Chicago, Ill., September 12, 1952.

The Development of Telephony in the United States

A. B. CLARK
FELLOW AIEE

Synopsis: The telephone was invented 24 years after the founding of the American Society of Civil Engineers and Architects, which society commemorated its Centennial in September 1952. The telephone was 8 years old when the American Institute of Electrical Engineers was founded. Since Bell's invention, the telephone business has grown tremendously and this growth has been greatly dependent on developments in science and engineering. This paper traces, and endeavors to give the significance of, the major developments. Brief mention is made of some developments still in the making and of some ideas as to the future potentialities of the business.

Table of Contents

Page
Organization of the Business........ 348
Transmission—The Struggle for Distance........ 349
Switching—The Early Days.......... 350
Co-ordination with Electric Power and Railway Systems................. 352
The Vacuum Tube and The Transistor 353
Transmission—Improving Service and Reducing Costs.................. 353
Switching—Developments Leading to a Nation-wide Automatic System.. 356
Plant Components Designed for Service at Minimum Cost............ 358
Some Services Associated with Telephony........................ 360
Radio........................... 361
Conclusion....................... 361

THE telephone was invented by Alexander Graham Bell in 1876. Two years later, in 1878, he wrote down a bold prediction. He wrote, in effect, that some day a grand system would be produced which would enable anyone anywhere to talk by telephone to anyone else, anywhere, regardless of distance.

The value of a connection by telephone cannot be measured by the number of words spoken for it brings people together in great intimacy so that understandings can be reached almost as well as in a direct meeting. The founders of the telephone business were, therefore, destined to create a service which would become of inestimable value to society. They had two great assets: a basic invention and an inspired charter. They had little else. The early telephone was crude indeed and there was almost no understanding of the problems of transmitting speech over distances of more than a few miles. Almost 40 years were to elapse before it would become possible to talk across the American continent.

The technical needs, which were of paramount importance, divided naturally into two classes:

1. To improve the telephone instruments and the lines so that speech reproduction would be loud and clear and remain so, even if transmitted over considerable distances. We now call this the "transmission problem."
2. To provide suitable switchboards and mechanisms whereby any two telephone instruments could be connected readily, to meet the desires of telephone customers. We now call this the "switching problem."

This paper contains four sections dealing with these problems, two of them dealing with developments in early days and two with developments in later days. Other sections deal with matters of importance both to transmission and switching.

Organization of the Business

An important problem was to organize the business so that as the technical needs were met, there could be developed the grand system envisioned by Alexander Graham Bell. Corporate structures were set up to this end. These eventually evolved into the present pattern of the Bell System, operating an integrated local and toll plant throughout the United States.

From the very beginning, a large amount of attention was given to the technical problems. In the parent company an experimental department was set up to study these matters. Then, within a few years, a patent department was also set up, since it was recognized that patents would become of large importance in the development of the business.

It was soon decided that the system needed its own manufacturing unit so that there would be a dependable source of supply, particularly of the items unique to the business. Also, it was recognized as highly desirable that there be a high degree of standardization of the different items. Accordingly, the Western Electric Company became a part of the infant system in 1882.

Guided by the engineering department of the parent company, working in close co-operation with the operating telephone companies, the manufacturing unit developed many useful things and, as time went on, became more and more effective in the field of development. Finally, the present setup evolved in which all of the research and development was consolidated in a separate company called Bell Telephone Laboratories, Inc., jointly owned by the parent company and the Western Electric Company. Included in Bell Telephone Laboratories are research and development for manufacture, formerly performed by the Western Electric Company, and also development engineering and research which had been carried out in the organization of the parent company. Bell Telephone Laboratories began operation in 1925. The absorption of the research and development activities of the parent company took place in 1934.

During the early days of the telephone business, there was intense competition so that in many towns there were two telephone companies sharing the business. This was unsatisfactory both to the telephone companies and to the customers. The problem was finally solved by acquisitions and mergers with broad regulation of service and rates of the single company which remained in each particular area. The Bell System, with 21 operating units, now serves about 82 per cent of the telephones, while an independent segment of the telephone industry, including more than 5,000 companies, serves the remainder. There are now 47,000,000 telephones in the United States.

In 1927, after public regulation of the telephone business had existed for about 20 years, Walter S. Gifford, then the president of the American Telephone and Telegraph Company, stated the basic principles of the business, which had evolved over the years, in his now famous Dallas speech before the National Association of Railroad and Utilities Commissioners. This policy was proclaimed to

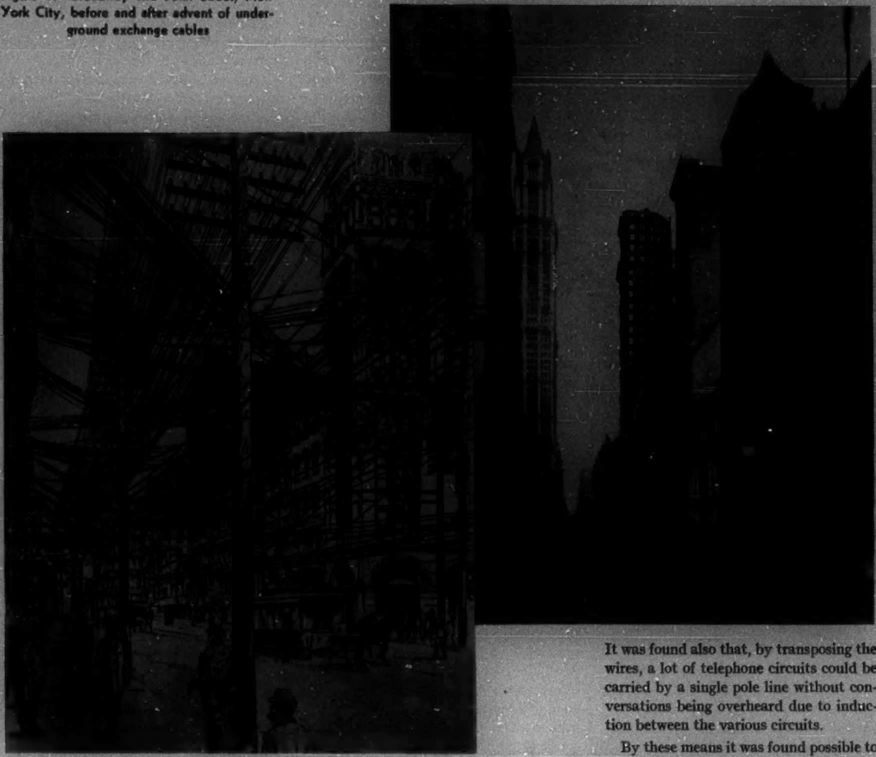
---

Paper 52-258, recommended by the AIEE Communication Division Committee and approved by the AIEE Technical Program Committee for presentation at the AIEE Centennial of Engineering—1952, Chicago, Ill., September 10–12, 1952. Manuscript submitted July 3, 1952; made available for printing July 25, 1952.

A. B. CLARK is with the Bell Telephone Laboratories, Inc., New York, N. Y.

To the considerable number of my associates who assisted, in various ways, in the preparation of this paper, grateful acknowledgment is made.

Figure 1. Broadway and John Street, New York City, before and after advent of underground exchange cables



the public in his annual report to the stockholders for the year 1928, from which the following quotation is taken: "Fair and just salaries and wages to the personnel, reasonable, regular dividends to the stockholders, together with an opportunity for them to make additional investments in the business from time to time on favorable terms, and further improvements in service or reduction in charges for service to telephone users so far as financial safety permits, are fundamental in the policy of the management to furnish the most telephone service and the best at the least cost."

These fundamental principles have guided the Bell System throughout the years. To the adoption of these principles must be credited a large measure of the success which has been achieved in the development of telephony in the United States.

We will now return to our narrative of progress in technical matters.

Transmission—The Struggle for Distance

By the end of the first decade after Bell's invention, there had been produced a practical form of telephone transmitter using carbon granules, the basic principle of present-day transmitters. This transmitter used energy obtained from a battery and thus made the electric waves corresponding to the voice very much stronger, the device being, in fact, an amplifier. Since the receiver was already fairly good, all that was then needed was a good line to carry conversations over distances of hundreds of miles.

An early step toward producing a good line was the development of hard-drawn copper, giving the tensile strength necessary for constructing a practical open-wire line. Shortly thereafter, experiments showed that by using a metallic circuit instead of a grounded circuit, a much quieter telephone line could be produced. It was found also that, by transposing the wires, a lot of telephone circuits could be carried by a single pole line without conversations being overheard due to induction between the various circuits.

By these means it was found possible to build a practical form of open-wire line to provide telephone service between New York and Boston. By using similar open-wire construction, and by making a number of other improvements, the range of telephone conversations was soon extended to distances of about 1,000 miles, so that it became possible to provide long-distance telephone service between the large eastern cities in the United States.

Meanwhile it was recognized that open-wire lines would not be practicable for providing the large number of telephone wires which would be needed in large cities. During the first decade after the invention of the telephone, practical means were developed whereby lead, heated to the point of plasticity, could be extruded over a compacted group of insulated conductors. This is the principle employed in most present-day telephone cables. Shortly thereafter, the idea of using dry paper as insulation for the conductors within the cables was introduced. This lowered the capacitance and made it possible to talk greater distances through these cables. It led ultimately to the general use of telephone cables for urban telephone service. Figure 1 graphically shows the difference the use of cables made to the appearance of a city street.

As telephone lines became longer and longer, a serious technical limitation became more and more apparent. This was the progressive weakening of the speech currents as the distance increased. With open-wire lines the speech remained clear as it became weaker. For cables, however, the speech became muffled as well as weaker. About all then known was that the effects became worse as the resistance of the line, or its capacitance, was increased. For very long lines, this demanded heavy-gauge copper wires. Open-wire lines naturally had low capacitance between wires but circuits in cables naturally had high capacitance. The step already mentioned of utilizing dry-paper insulation for cables was dictated by the desire for a low-capacitance line. Having taken this step, it was not practical to lower the capacitance further. This left the distance range of cables for satisfactory conversations seriously limited.

The "phantom" circuit was introduced around 1900. It reduced, but alone could not check, the rising cost of the lines as their length became greater.

The phantom principle made it possible to obtain three telephone circuits from two pairs of wires which previously had yielded only two telephone circuits. An additional circuit, called the phantom was obtained by using, as one conductor, the two wires of one pair in parallel and, as the other conductor, the two wires of the other pair. To make the principle practical required the development of efficient balanced transformers whereby connections could be made to the phantom circuit without disturbing conversations carried on over the pairs. It was also necessary to develop a system of transpositions or twists in the wires of both the phantom and the physical circuits to keep crosstalk between the three circuits, and other circuits carried on the same line or within the same cable, within tolerable bounds. These were no mean achievements—it is small wonder that it took something like 20 years to reduce the invention of the phantom principle to a practical realization. The phantom principle then began to play an important part in telephony and soon the open-wire lines and new intercity cables were almost completely phantomed.

Meanwhile theoretical inquiries, aimed at better understanding of the attenuation process, had been proceeding. Mathematical formulas were finally worked out which explained the phenomena both qualitatively and quantitatively. These inquiries led, in 1899, to the invention of coil loading. To "load" a line, lumps of inductance called "loading coils" are inserted at regular and frequent intervals in series with the wires of the telephone line. The formulas had already shown that increasing the series inductance of a telephone line would greatly reduce its attenuation. The invention constituted a practical means for obtaining the result. Loading made it possible to double the length of open-wire line over which conversations could be satisfactorily carried on and to improve transmission through cable circuits even more. Not only was the attenuation reduced but also the distortion which previously had caused the muffled sounds.

The practical application of loading took place concurrently with the application of the phantom principle. It then became practicable and not prohibitively expensive to extend the useful range of open-wire lines to approximately 2,000 miles. Loading and phantoming also made it practicable to connect Washington, D.C. with New York City by underground cable, which was soon extended to Boston. The whole project was completed in 1913. Some years before, during the inauguration of President Taft, a severe sleet storm had broken down the open-wire lines and isolated Washington from the rest of the country. The construction of the underground cable effectively prevented anything like this from ever happening again.

These extensions of open wire and cable remained the extreme limits of practical telephony until the dawn of the electronic era, in which the vacuum tube made it possible to stretch the range of telephony to unlimited distances by wire and radio. From that time on the major telephone transmission jobs have been to improve service and reduce costs. There remained, however, many other jobs of great technical interest and importance. Before considering transmission further, however, we must catch up with the happenings in the switching field.

Switching—The Early Days

The early telephone sets were mounted on a wall and included a separate transmitter and receiver for talking and listening, and a bell for signaling the customer to the telephone. The set also included a magneto generator, cranked by hand, for signaling the telephone operator, and a box containing batteries which supplied talking power to the transmitter. Although considerably changed in appearance, a few telephone sets in use even today include all of these features.

At first, these telephones were leased to customers who built their own lines and used them for their own private service. Soon, however, it became evident that the usefulness of the telephone would be greatly increased if means were provided whereby different telephones could be connected on demand. This led to the creation of centralized telephone switchboards, originally operated by boys. These boys made connection from one single-wire telephone line to another as best they could, considering that both the lines and the operating rooms were so noisy that the faint voices coming from customers' telephone lines could barely be understood.

However, the setup was soon improved. New or modified switchboards were provided in which the single-wire grounded lines were replaced by metallic circuits. This quieted the connections, so that customers and operators could hear better. Female operators replaced the boys. A major improvement was made by providing a storage battery at the central office which supplied electric current for powering the transmitters of the customers' telephones. This same common battery also supplied power for signaling an operator in the central office when the telephone customer removed the receiver from its hook and replaced it at the end of his conversation.

Figure 2 shows, in the upper lefthand picture, one of the first telephone central offices and below this a modern manual switchboard. At the right is a view of some of the switching equipment in a modern crossbar office.

Telephone lines connecting one community with another soon began to be built. Thus began toll telephone service in which telephone customers served by an exchange at one point were connected over toll lines to customers served by another exchange some distance away. This involved more complications than local calls and also required the telephone operators to take on another job, that of timing and ticketing the calls.

Problems became more complicated when really long-distance calls began to be handled. The long-distance lines were so expensive that it was unthinkable to provide enough lines to handle the service on a no-delay basis. It was also necessary to minimize transmission losses, wherever

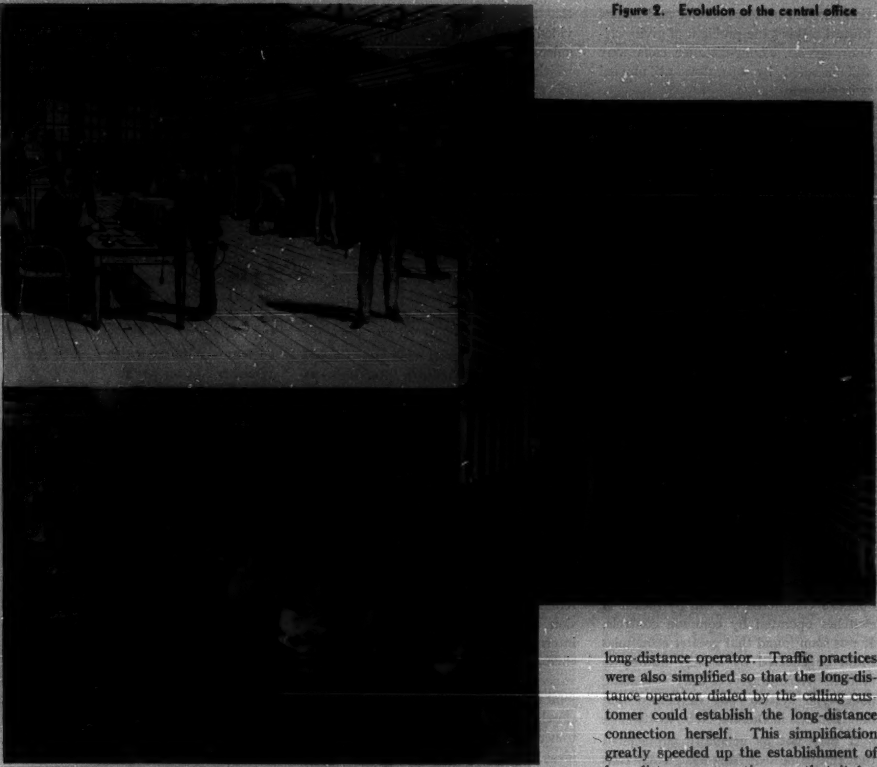


Figure 2. Evolution of the central office

practicable at the terminating ends of the connections. These things led to a procedure for handling long-distance calls which now appears clumsy indeed. The procedure was as follows:

When a telephone customer wished to place a long-distance call, his local operator connected him to a long-distance operator about as she would make connection to a local telephone. When the long-distance operator answered, the calling customer gave the name and address (or telephone number if he knew it) of the person he wished to reach. The long-distance operator wrote this on a ticket and then pulled down the connection to the calling customer and established a new connection to him using a trunk with large-gauge conductors and also, to improve transmission, a higher-voltage central office battery. Another operator then took the ticket and noted on it how the call should be routed as well as the initial-period charge. Then, still another operator proceeded to set up a connection to the called customer's telephone. After getting the called customer on the line, she then called back the calling customer, using the connection already established by the first operator.

In due course, it was found desirable to use the higher-voltage central office batteries for both local and long-distance calls. It was not, however, until after the vacuum tube had unshackled progress in the transmission art that it became possible to eliminate the whole clumsy procedure. Long-distance circuits then could be provided in quantities ample to handle most calls without delay. The amplification provided in the long-distance lines, coupled with transmission improvements in telephone instruments, made it no longer necessary to use trunks with large-gauge conductors in the connection between the calling customer and the long-distance operator. Traffic practices were also simplified so that the long-distance operator dialed by the calling customer could establish the long-distance connection herself. This simplification greatly speeded up the establishment of long-distance connections so that it became the normal procedure for the calling customer to remain at his telephone until the distant party was reached. Also, the way was paved for the automatic handling of long-distance calls which was to follow later.

Meanwhile, a number of inventors had been at work on automatic switching systems in which a telephone customer would be able to call another without the aid of an operator. One of these is now called the "step-by-step" system. The first step-by-step system was placed in service in 1892 at La Porte, Ind., where it served about 75 customers. With this system, switches in the central office respond directly to each digit as it is dialed by the calling customer. Thus, the connection is extended step by step to the called telephone.

This system was improved from time to time. By 1908 step-by-step exchanges were in use in about 70 cities and towns in the United States. Automatic operation speeded up local service and helped toward solving the problem of giving 24-hour service. Long-distance calls were still handled by operators at manual switchboards.

After many experiments with different forms of automatic switching systems, the Bell System undertook the development of an automatic switching system which would be particularly suited to multi-exchange service in large cities. The system used a new form of switch called the "panel" type, and embodied a new principle of switching circuitry called "common control." The common control circuits are separate from the switches which establish the customer's connection. They receive the pulses dialed by the customer and determine the routing, after which they cause switches to be operated to make the desired connection. Then they drop off and become available for establishing other connections. Common control switching systems possess great versatility and become more and more advantageous as the switching job to be performed becomes more and more complicated.

A panel system for "semimechanical" operation was installed in 1915. The customer gave his number by voice and an operator completed the call by means of switches operated by common controls. It was soon found that greater economies could be effected by a fully automatic system. Accordingly, a fully automatic panel-type system was developed and first installed in Omaha, Neb., in 1921.

It may not be generally appreciated that telephone switching has greatly contributed in reducing the cost of telephone service. This is because probability is invoked so that much less plant is provided than would be needed to permit all customers to carry on conversations at the same time.

It is very important to strike a nice balance, so that enough plant will be provided to give adequate service, but not more than enough, since this would make the plant unduly expensive. This has required the development of measuring devices to check the loads carried by different portions of the plant. It has also required the development of a new kind of applied mathematics known as "trunking theory." The first formal papers on trunking theory began to appear soon after the turn of the century and since that time mathematicians both in this country and abroad have devoted considerable attention to this subject. Some idea of its importance can be gained by citing a few illustrative figures. Today in large metropolitan areas it is estimated that the number of trunks actually provided for connections between central offices is less than 1 per cent of the number which would be required for connecting all customers who conceivably might wish to talk at the same time. It is also estimated that in a large automatic central office the over-all quantity of switching equipment is only about 10 per cent of that which would be required to permit all of the customers served by the office to converse simultaneously. Some of the early inventors of automatic switching systems did not appreciate this and were, therefore, thwarted in their efforts to produce a practical system.

Trunking theory is, however, only one facet of switching development. Over the years, the circuitry of switching systems has become highly refined. Starting as an art, the principles of design of switching circuits are now assuming the aspects of a science. These principles were recently incorporated in a textbook for study in universities. The subject has become of wide interest not only in telephone switching but also in the electrical computer and many other industries.

We will consider switching further after catching up on other developments. Before proceeding, however, the following remarks may be helpful. The transmission and switching fields were closely interrelated, advances in one making possible advances in the other. Large advances were made, resulting from the joint efforts of telephone operating people closely attuned to the needs of their customers, of others who developed new ideas and instrumentalities for improvements, and of still others who manufactured equipment and continuously improved it. Thus, the service was greatly improved over the years. The improvements, coupled with an increasing number of telephones which each subscriber could reach, greatly increased the value of the service and caused a rapid growth in the business.

Co-ordination with Electric Power and Railway Systems

The electric power industry grew up with the telephone industry. By 1878, arc lighting systems were in operation, and in 1882 incandescent electric lights and trolley cars were in service. Power lines located near the ground return telephone lines of that day made the telephones very noisy. A big step was taken in reducing the noise when the ground return telephone lines were replaced with balanced metallic circuits. The effects of the noise were further reduced by the introduction of improved telephone transmitters, which made the voice currents stronger.

However, as power systems expanded and telephone lines became longer, difficulties with noise increased. Also, induction from power lines and electrified railways tended to cause false ringing of telephone bells and disturbances in telegraph circuits and, under some conditions, to create hazards to telephone plant and telephone employees and customers. There was also increasing danger of serious troubles due to accidental contacts between power and telephone wires.

It became evident that the problems could be most effectively solved by cooperation between engineers and executives of the two industries. Joint committees were formed which led to general agreement on guiding principles for action whenever there was need to place power and telephone lines in proximity. It was agreed that:

1. The two industries would endeavor to operate their plants in accordance with principles and practices designed to reduce the probability of hazardous contacts or inductive disturbances.

2. In specific cases of trouble the "best engineering solution" should be sought. This meant that corrective measures would be devised for application in either or both systems which would result in the lowest over-all cost.

3. Strong endeavors would be made to obtain as complete an understanding as possible of all of the technical factors involved in problems of co-ordinating power and telephone lines. Only by such an understanding could effective and economical remedial measures be worked out.

This agreement led to an extensive program of co-operative tests and theoretical studies in which telephone engineers and engineers of the power operating companies and electrical manufacturers participated. Joint work of a similar nature was also initiated with engineers concerned with railway electrification projects.

The result of this co-operative approach has been to make available to all concerned a great wealth of technical information together with a set of guiding principles for the co-ordination of electric power and communication systems. By following these principles, particularly in the advance planning of power and telephone plant construction, it has been possible for the operating companies to bring the co-ordination problem under control. Even today, however, new developments in both the power and telephone industries require constant scrutiny from this point of view and there is still a need for engineers skilled in this art to make certain that adequate co-ordination of the two utilities systems is maintained.

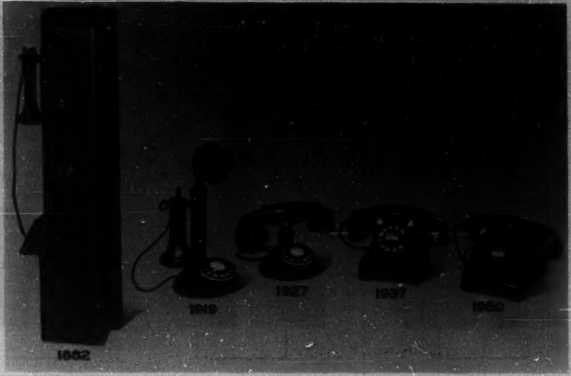


Figure 3. Progressive changes in the telephone instrument

The Vacuum Tube and the Transistor

Early in the 20th century, Lee DeForest made a great electronic invention, the amplifying vacuum tube. Bell System engineers, recognizing the potentialities, acquired rights to use the invention and inaugurated a program of research to discover methods of improving it. They soon came to the realization that a major requirement for good performance was the creation of an extremely high degree of vacuum within the tube. The resulting high vacuum tubes more than fulfilled their expectations. The solution of the telephone repeater amplification problem was at hand.

Using telephone repeaters in which the amplification was supplied by vacuum tubes, telephone service between the east and west coasts of the United States was inaugurated in January 1915. This was followed in the same year by demonstrations that, by means of vacuum tubes, voices could be carried by radiotelephony to great distances—across the Atlantic to the Eiffel Tower in Paris and across the Pacific to Hawaii. Thus it became evident that for telephony a means had been found for the conquest of distance.

The vacuum tube also supplied, for the first time, a means for precise measurement of the weak currents and the frequencies characteristic of telephony. Vacuum-tube oscillators of variable frequency could be used to supply the excitation. Other vacuum tubes amplified the currents to be measured and rectified them so that they could be conveniently measured with existing meters. The ability to make precision measurements established a firm foundation for the transmission art. It also made it possible to build on this foundation the repeatered lines and carrier systems which revolutionized long-distance telephony.

The vacuum tube also made it possible to create two great broadcasting industries—radio or, better stated, audio broadcasting, and television or video broadcasting. Telephone engineers played a large part in the early development of these industries. They are both very important to the telephone business since they constitute a large market for transmission facilities which can be advantageously supplied along with telephone circuits.

Four years ago, Bell Telephone Laboratories announced the invention of a new electronic device which they named the "transistor." A transistor is a 3-element device which depends for its functioning on a semiconductor, such as germanium. Variation of the current in the circuit of a "control" element, corresponding somewhat to the grid of a vacuum tube, causes corresponding changes in the voltage in the circuit of a "collector" element, which corresponds somewhat to the plate circuit of a vacuum tube.

For many applications, transistors have capabilities comparable with vacuum tubes. They will do many jobs more economically because they require no source of energy for heating a cathode and because they can be operated with lower voltages for the same amplification. They also have the advantage of inherently smaller size than vacuum tubes, which is of great importance in some applications.

Transistors should find large application in switching, particularly in common control circuitry. In switching, electronic tubes have not had anything like the profound effect which they have had in transmission, although many vacuum tubes and cold cathode tubes are used today in switching systems. Among engineers there is today considerable speculation concerning the extent to which transistors will revolutionize the switching systems of the future.

The transistor will undoubtedly find other large applications in telephony as well as in other fields. In telephony, transistors will make it economical to use amplification much more widely in the local plant, rural, and short-haul toll fields. Other fields covering, for instance, hearing aids, electric computers, radio and television sets, and many military applications offer even greater potentialities.

Transmission—Improving Service and Reducing Costs

Over the years, the instruments provided for telephone customers have been greatly improved. The improved designs have resulted from:

1. Careful studies of customers' needs and the results of operating experience.
2. Research in different fields including acoustics and materials.
3. The work of designers familiar with field experience, the possibilities of research, and the requirements for economical manufacture.
4. The efforts of a manufacturer well schooled in large-scale quality production.

The modern telephone set possesses the good appearance characteristic of modern design and is an extremely efficient and convenient instrument. Conversations have become practically effortless except under very adverse environmental or circuit conditions. These instruments have made it possible to reduce the sizes of wires in local cables so that many of them are now about as small as can be manipulated conveniently by human hands. Figure 3 shows the new instrument compared with some of the models of earlier years.

When vacuum tubes made unlimited and practically distortionless amplification available for application to telephone lines, a number of transmission defects became apparent. Electrical echoes were noted which affected both the talker and the listener. The talker was disturbed by hearing echoes of his speech and the listener by hearing hollow sounds and in some cases actual echoes of the speech from the talker at the distant end of the

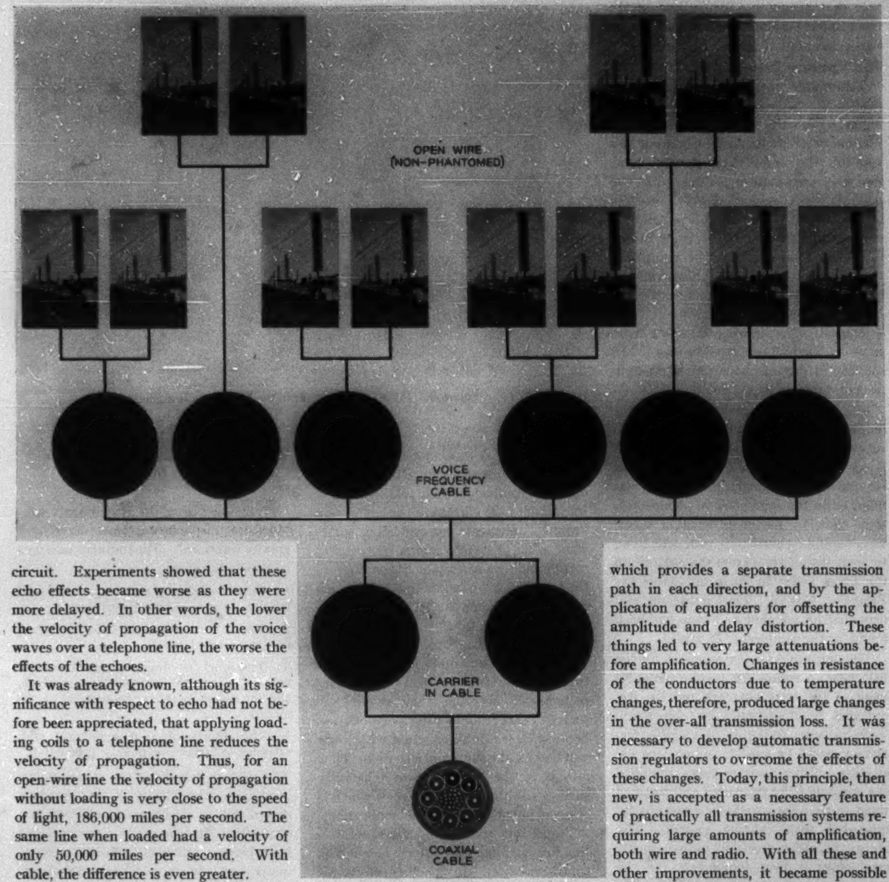


Figure 4. Trend in outside wire plant

circuit. Experiments showed that these echo effects became worse as they were more delayed. In other words, the lower the velocity of propagation of the voice waves over a telephone line, the worse the effects of the echoes.

It was already known, although its significance with respect to echo had not before been appreciated, that applying loading coils to a telephone line reduces the velocity of propagation. Thus, for an open-wire line the velocity of propagation without loading is very close to the speed of light, 186,000 miles per second. The same line when loaded had a velocity of only 50,000 miles per second. With cable, the difference is even greater.

Other transmission deficiencies of loaded open-wire lines were severe distortion both with respect to attenuation and transmission delay, and great transmission loss variations due to variable leakage. Thus, it soon became evident that the repeatered open-wire lines of the future should not be loaded. The loading coils were, therefore, removed from the open-wire lines and the increased attenuation overcome by doubling the number of repeaters. As predicted, the transmission became louder and clearer and at the same time less affected by rain.

Meanwhile the application of repeaters to cable circuits proceeded. It was soon evident that large-gauge conductors would no longer be necessary, so that today the size of wire most commonly used in toll cables is number 19 American wire gauge weighing 20 pounds per wire mile. For comparison, the largest wires used in the underground cables between Boston and Washington were number 10 American wire gauge weighing 160 pounds per wire mile.

In cable circuits, some adverse effects from loading, similar to those already noted for open wire, were also experienced. The effects were even more serious because of their higher capacitance. These adverse effects were minimized by making a large reduction in the inductance of the loading coils, by using 4-wire operation which provides a separate transmission path in each direction, and by the application of equalizers for offsetting the amplitude and delay distortion. These things led to very large attenuations before amplification. Changes in resistance of the conductors due to temperature changes, therefore, produced large changes in the over-all transmission loss. It was necessary to develop automatic transmission regulators to overcome the effects of these changes. Today, this principle, then new, is accepted as a necessary feature of practically all transmission systems requiring large amounts of amplification, both wire and radio. With all these and other improvements, it became possible to extend the range of satisfactory transmission through cables to distances in excess of 1,000 miles.

Meanwhile, however, research was creating a powerful new transmission method to which the name "carrier system" was applied. A carrier system uses a number of different frequency bands separated by electrical filters to carry a number of different messages simultaneously over one transmission path. The first application of this principle was made on an open-wire line in 1918. Perfecting this system required the surmounting of many difficulties. Today a single open-wire pair can provide by this method as many as 16 telephone circuits of excellent quality.

In the early 1930's the carrier and

vacuum-tube arts had been perfected to the point where application of carrier systems to cable circuits became practicable. This was greatly aided by the invention of the negative feedback principle. When negative feedback is applied to a vacuum-tube amplifier, the already low distortion is reduced practically to the vanishing point while the over-all amplification becomes highly stable, almost completely immune from variation due to tube aging, battery noise, and voltage variation. In addition, great progress had been made in the understanding of methods for the control of noise and cross-talk, which contributed importantly.

The carrier principle applied to cables makes it possible to obtain from two 19-gauge nonloaded pairs, 12 telephone circuits of excellent quality, so good, in fact, that the length of the cables can be extended to thousands of miles without appreciable distortion. In common with open-wire carrier, these carrier-in-cable circuits have the advantage of high velocity so that where they are used, echo effects are no longer a matter of serious concern.

Carrier in cable was followed shortly by an even more radical development. This was the introduction into the cables of a new type of conductor just a little larger in diameter than a lead pencil (3/8 inch), called a "coaxial conductor." This consisted of a copper tube with a wire running down its axis and insulated from it by washers a little less than an inch apart. By applying feedback amplifiers, spaced about 8 miles apart, to these coaxials, it became possible to transmit a band of frequencies nearly 3,000,000 cycles wide. This is today yielding 600 two-way telephone circuits from a pair of coaxial tubes. Alternatively, television programs can be transmitted.

Using the same coaxial conductors, development is well along, and field trials have begun of an improved system in which the band of frequencies transmitted over one coaxial will be extended to nearly 8,000,000 cycles. With this system the coaxial conductors will remain the same, but amplifiers, greatly improved in design and with improved arrangements for equalization and regulation, will be placed at 4-mile intervals instead of the present 8-mile intervals. This system is expected to be capable of transmitting as many as 1,800 telephone circuits over one pair of coaxials. Alternatively, one of the coaxials might be used for the transmission of high-definition television, or perhaps to obtain a 4-megacycle channel for present-day television plus about 600 one-way telephone channels.

Figure 4 shows how these transmission developments have resulted in great compacting of the outside wire plant. Two coaxials in a cable will carry as many conversations (1,500+) as two cables equipped for carrier, six voice-frequency cables, or 12 250-wire nonphantomed open-wire lines.

Very recently, microwave repeatered radio has been developed to the point of practical application, as another method for obtaining long-distance telephone circuits in large quantities, as well as broad-frequency ranges for television transmission. With this system, radio relay towers are spaced approximately 30 miles apart. Horn-shaped antennas are placed on the tops of these towers with line-of-sight transmission paths between the antennas. Such paths are obtained by utilizing hills to the greatest extent practicable so that tower heights will not be excessive. Figure 5 shows a typical microwave relay station.

The construction of a microwave radio relay system spanning the United States was completed in 1951 and coast-to-coast television network transmission inaugurated in September of that year. This microwave system is capable of transmitting six broad-band channels in each direction. Each channel can transmit a television program or, alternatively, as many as 600 telephone channels. Thus, the over-all transmission capacity of such a system is comparable with that of a cable containing eight coaxials. Radio and coaxials are natural complements to each other with respect to suitability for different terrains and to vulnerability against different causes of service interruption. Both are in keeping with the large numbers of long-haul telephone circuits now provided on major toll routes, as shown in Figure 6.

These broad-band carrier systems, open-wire, wire cable, coaxial cable, or microwave radio, have many transmission similarities. All of them use the same 12-channel equipment as the basic building block for obtaining individual telephone channels, and these building blocks comprise the bulk of the terminal apparatus for all of the systems. For all of the systems the voice waves are transmitted from point to point at a velocity approaching the speed of light. As already noted, this minimizes trouble from electrical echoes. Echoes are further minimized by the fact that for all of the systems the telephone circuits are operated on a 4-wire basis. The vacuum tube has caused attenuation to be no longer a matter of concern with any of these systems. Problems still remain, however, in keeping the circuits quiet and free from outside disturbance, while at the same time holding the over-all transmission loss at constant value.

All of these systems achieve low cost by making it possible for a single telephone route to carry large numbers of circuits. Thus, large economies are obtained by dividing the common elements of cost, including right-of-way costs, costs for cable sheath and cable laying, the cost of radio towers, and the cost of maintenance between large numbers of telephone circuits. The coaxial and the radio relay systems carry this process to the point where less than a thousandth of the common costs are borne by a single telephone circuit.

For the coaxial and radio relay systems it will of course be evident that there is much economic advantage to be gained by adding a number of television channels to the telephone channels carried over a single route. This is because the transmission of a television image requires a frequency band wide enough to furnish hundreds of 2-way telephone circuits. Thus, with only a few television channels added to a route, the total frequency range provided for television service becomes comparable to that for telephone service. The common costs of the route are shared by the two services, with resultant cost reduction benefits for each.

Perhaps some day the quantity of transmission facilities demanded will outgrow the capacity even of these coaxial and microwave radio systems. Against this possibility, research has been proceeding for many years on transmission through pipes called "wave guides." A wave guide provides a private ether path over which a frequency band a billion cycles wide can be efficiently transmitted. Exploiting the possibilities of wave guides

Figure 5. A typical microwave relay station

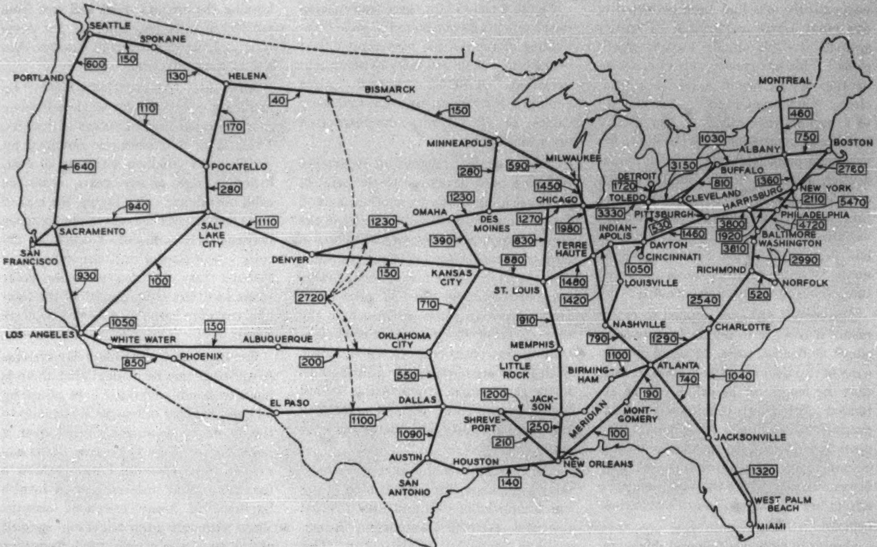

SEATTLE
SPOKANE
600
150
130
HELENA
PORTLAND
40
BISMARCK
150
110
170
MINNEAPOLIS
640
POCATELLO
280
590
MILWAUKEE
280
1230
1450
940
SACRAMENTO
SALT LAKE CITY
1110
1230
OMAHA
DES MOINES
1270
CHICAGO
SAN FRANCISCO
390
830
1980
DENVER
150
KANSAS CITY
880
TERRE HAUTE
930
100
2720
710
ST. LOUIS
1480
1420
1050
910
150
LOS ANGELES
WHITE WATER
ALBUQUERQUE
OKLAHOMA CITY
MEMPHIS
PHOENIX
850
200
LITTLE ROCK
550
1200
EL PASO
DALLAS
JACKSON
1100
SHREVEPORT
250
1090
210
AUSTIN
HOUSTON
140
SAN ANTONIO
MONTREAL
460
1030
750
BOSTON
ALBANY
3150
BUFFALO
2760
DETROIT
1720
810
1360
NEW YORK
TOLEDO
CLEVELAND
HARRISBURG
2110
5470
3330
PITTSBURGH
PHILADELPHIA
INDIANAPOLIS
530
1460
3800
4720
DAYTON
1920
BALTIMORE
3810
WASHINGTON
CINCINNATI
2990
1050
RICHMOND
LOUISVILLE
NORFOLK
2540
520
NASHVILLE
CHARLOTTE
790
1290
1100
ATLANTA
BIRMINGHAM
740
1040
190
MERIDIAN
MONTGOMERY
100
NEW ORLEANS
JACKSONVILLE
1320
WEST PALM BEACH
MIAMI

Figure 6. Long-haul telephone circuits on major toll routes

for long-distance transmission will require the creation of amplifiers capable of efficient operation at frequencies approaching a hundred billion cycles. Although the wave guide is not yet a candidate for long-distance transmission, it has found many important applications in the local circuitry of microwave radio systems as well as in radars and other military devices.

Although carrier systems, particularly when supplied in large quantities on a single route, make it possible to effect large reductions in line cost, these tend to be offset, when the circuits are short, by the cost of the carrier terminal apparatus. Until a few years ago, carrier usually could not be proved in economically against fine-wire cable circuits worked on a voice-frequency basis for distances much shorter than 75 miles. A new 12-channel cable carrier system has recently been developed, however, for which the cost of the terminals is low enough so that the system is suitable for lengths as short as 20 miles. A somewhat similar system for application to open-wire lines is now going into service.

In rural areas, economic application of carrier is difficult because the number of circuits needed is so small that the common costs for the line remain high even when divided among the several circuits furnished. One way to hold down costs is to use the same wires for transmission of power and for telephone service. To exploit this possibility, a rural power-line carrier system was developed a few years ago and a number of such systems are now in service. Applications of this system will, no doubt, expand as it is further improved and reduced in cost. The transistor offers further possibilities for reducing the cost of rural telephone service.

A significant forward step in submarine telephone cable development was taken in April 1950 when a pair of repeatered submarine cables was laid between Key West, Fla., and Havana, Cuba, a distance of 140 geographical miles. Each one of these cables has built into it three 1-way vacuum-tube amplifiers. These are flexibly enclosed in watertight housings, protected by armor wires, so that in over-all appearance they simply look like enlargements of the cable about 20 feet in length. The pair of cables provides a total of 24 telephone circuits. It is quite possible that these may be the forerunners of other repeatered submarine cables for giving telephone service between America and Europe.

Switching—Developments Leading to a Nation-wide Automatic System

In the early 1920's automatic telephone switching had become well established so that by 1925 about 10 per cent of the telephones in the United States were of the dial type.* The step-by-step system had become established in the smaller towns, and various improvements introduced into the system made it possible to extend its application into larger communities. The panel-type common control system had become established in metropolitan areas. In these areas the common control principle was beginning to prove its worth in handling the trunking problems which became increasingly complicated as the cities grew larger and the telephone development expanded.

Then, as transmission became better and as switching systems were further improved, automatic switching began to be applied more and more to facilitate the setting up of connections from a customer to a larger and larger number of telephones. This developed in two ways:

1. Telephone operators began to dial over toll lines into automatic systems in nearby towns which made connection to the desired telephone without requiring the services of another operator. The dial pulses were transmitted by d-c methods already developed for telegraph. The long-distance operator ticketed the calls in the usual way.
2. The switching systems, particularly in the larger cities, were expanded so that telephone customers could dial telephones at greater distances, thus greatly increasing

---

* Today, over 70 per cent are dial.

the number of telephones which they could reach without an operator. This, like item 1, required the development of tandem switching centers since it would not have been economical to provide direct trunks to the more distant exchanges. It also required the development of means for automatic registration of the charges for the calls.

The development of automatic methods for recording the data required for the proper charging of telephone calls was destined to become of great importance in the future extension of automatic switching. A brief account will therefore be given not only of the registering system which was first developed but also of progress which was made in the years which followed, toward methods suitable for more extended application.

The automatic registration of telephone calls was first accomplished by means of message registers (already in use in manual service). A message register located in the telephone central office is associated with each customer's telephone. Apparatus was added to the automatic switching systems so that these registers would automatically record the number of "message units" made by each customer. In the most elaborate arrangement, which was applied to some of the larger metropolitan areas, means were provided so that registers would be operated more than once, if required, on a single call, the number of operations depending on the distance and the duration of the call. Each customer was billed on a bulk basis for all of the message units totaled on his register. In view of the fact that detailed billing of individual calls is not possible, it has seemed desirable in the United States to limit this method of message registration to the shorter-haul calls.

A later method developed for charging is known as "automatic ticketing." This method was first introduced in Belgium where it is still in use, although its use has not spread to other points in Europe. In this country a somewhat similar system was placed in service in California in 1944. This system is associated with step-by-step local switching equipment. It automatically prints for each call a ticket similar to that prepared by the operator in manual operation. These installations are somewhat expensive since a considerable amount of gear auxiliary to the printers must be added in each office. It is unlikely that there will be much further extension of use of this method.

A greatly improved type of message registration system was later developed and first installed for service in Media, near Philadelphia, Pa., in 1948. In this system the information required for the billing of calls is recorded as holes in a multitrack paper tape. These tapes are processed in accounting centers by automatic methods which readily lend themselves to automatic preparation of toll statements or bills. In these bills, calls may be detailed to any extent considered desirable. This system is called "Automatic Message Accounting" (usually abbreviated to AMA). AMA is adequate for handling all calls dialed by customers regardless of distance or duration. The use of this system is being rapidly extended. In Figure 7 an AMA accounting center is illustrated.

Meanwhile, as the circuitry of common control systems was further developed, it became possible to produce a much improved switching system of this type, called the crossbar system. In this system, the panel switches are replaced by much smaller crossbar switches. This replacement was made practicable by the refinement of a principle in switching known as "marker" operation. The crossbar switches are superior because they operate faster and because they make connections by means of relay-type springs. These are equipped with precious metal contacts, assuring reliability and complete freedom from noises of microphonic character sometimes experienced with switches using base metal contacts.

Good results were obtained from the crossbar system and its success led to the development of other common control automatic switching systems using switches of the same type. One of these was a crossbar system particularly adapted to the automatic switching of long-distance circuits. With this system, connections between long-distance circuits are made on a 4-wire basis. This eliminates any possibility that transmission will be impaired by the switch. The first system of this type was placed in service in Philadelphia in 1943. During World War II, this system received a severe test, which it met very successfully, in handling larger than anticipated amounts of long-distance traffic.

At the same time, a big expansion of long-distance service had taken place, brought about by a combination of favorable factors. The development of fine-wire cables and carrier systems had made it possible to reduce long-distance rates. The transmission performance of long-distance circuits had been greatly improved. Telephone connections, some with as many as five circuit links in tandem, were being regularly and successfully established by telephone operators. The customers were pleased by the promptness of the establishment of long-distance connections.

All of these things, coupled with the success of the installation at Philadelphia, led to studies of the service and operating results which might be expected from a nation-wide extension of automatic switching. The conclusion was reached that this would be a desirable objective. Accordingly the outline of a national plan was prepared and adopted by the telephone companies. It will embrace all of the United States and extend into Canada and perhaps into Mexico.

While the Bell System necessarily took the lead in this planning, the details are being arranged so as to include the tele-

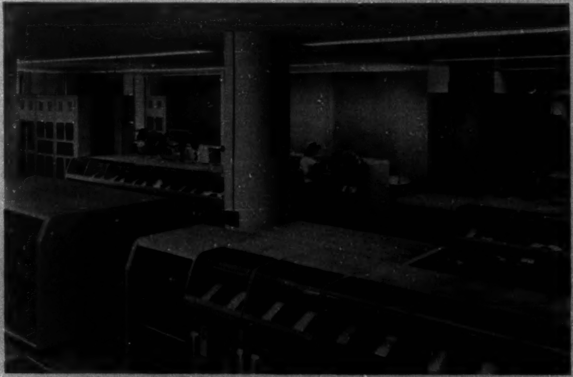


Figure 7. Automatic message accounting center at Philadelphia, Pa.

phone users served by independent telephone companies. The independent companies are being kept fully informed as the plan develops and are participating, as the development of their own plants makes it practicable and desirable, in extending its benefits to their own customers.

In accordance with this plan, all of the telephones will be assigned a long-distance telephone number consisting of ten digits (11 where a party letter is involved) of which the first three digits designate the area in which the telephone is located, there being about 100 such areas. The other seven (or eight) digits designate individual telephones in each area. It is not necessary to use more digits than this for calls between telephones in one area, which may embrace a whole state.

At somewhat less than 100 strategic points in the countrywide network of long-distance circuits there will be placed automatic long-distance switching systems. These will be similar to the one at Philadelphia, already mentioned, but with a number of added features which will give them greater versatility and higher intelligence so that their performance will become, to the uninitiated, almost magical. When the 10-digit designation of any telephone (seven for intra-area calls) is dialed into one of these robots, it will immediately and automatically start the establishment of a connection to this telephone. On its way this connection may pass through one or more additional robots. These robots attempt to make connection first over the most direct routes. If a circuit over a direct route is not available they will then try other routes in accordance with a systematic and prearranged plan until connection is finally made to the called telephone through the local automatic central office from which it regularly receives its service.

It will, of course, be evident that in a simplified discussion of this plan it is not practicable to cover all of the important features. To mention one, there is the obvious necessity for providing a-c signaling systems for transmitting the telephone numbers at high speed over the circuits, along with other signals required for various auxiliary purposes. Another is the necessity for providing auxiliary automatic toll switching offices at a large number of points where connection is made between the local and long distance networks. These systems require less intelligence than the long-distance robots already mentioned and also do not require the feature of 4-wire switching.

In the first stage of the plan, the telephone customers will place their long-distance calls with the operator as at present. She will dial the long-distance number of the called telephone into the network and the connection to this telephone will then be automatically established. Today, operators are dialing about one-third of their toll calls and 18 long-distance switching robots are already in service. Temporarily, the operators are using routing codes instead of the long-distance telephone numbers. Also the robots do not yet include all of the features required for the nation-wide plan. The first robot fully equipped in accordance with the plan will go into service in 1953 and other installations will follow rapidly. Then in due course the present robots will have incorporated into them the added features required for the plan.

When the plan is completed, the arrangement of the long-distance network will be such that customers, if properly connected to it, will be able to dial their own long-distance calls. This will involve many problems, such as the extent to which customers will wish to dial such calls, their accuracy of dialing, the treatment of wrong numbers, and methods of obtaining telephone numbers in distant cities. Since the best way to develop these questions is a trial, arrangements were made to inaugurate such a trial at Englewood, N. J., last fall and the trial is still in progress. Facilities have been provided so that 10,000 Englewood customers are able to dial directly any one of about 11,000,000 telephones in ten metropolitan areas including Boston, New York, Chicago, and San Francisco. The results to date are in favor of nation-wide extension of this service.

At Englewood the telephone customers are served by the latest type of crossbar switching system which was designed with the requirements in view of nation-wide customer dialing. It was a simple matter to arrange the system at Englewood to receive the 10-digit long-distance telephone numbers from the calling customers, extend the connection to the long-distance network, and then pass along the telephone number. The system was also already equipped for AMA operation so that the data required for charging the calls could be registered automatically, although some added features were required for the nation-wide application.

It will be evident that for universal application of nation-wide customer dialing it would be necessary that all of the local central offices, in combination with the auxiliary toll switching centers which provide access to the long distance network, be arranged to provide these features. In some areas, this would not be easy. However, one can hope that some day it will become practical to do these things. All switching systems would then play a role in the nation-wide symphony. This would indeed be a wonderful culmination of automatic switching development.

Fortunately, it is not necessary for the success of the nation-wide system that customer dialing of long-distance calls be provided immediately in all areas. Customer dialing of long-distance calls can be practiced in some areas, and in other areas the calls can be handled by operators, without doing violence to the nation-wide plan. In each area there is therefore a choice which can be made between operator or customer dialing. Choice as to procedure will depend on the cost of equipping local central offices for customer dialing, providing the required access to the nation-wide network, and also providing facilities for automatically charging the calls; and on the desires of customers for nation-wide dialing.

It is of interest to note that with all of this increase in automatic switching the number of telephone operators has not decreased. This is because expansion of the business has accompanied increased utilization of automatic methods. In the future there will still be a large amount of work for operators, even if it is assumed that all of the local offices have been arranged to permit customer dialing of long-distance calls. The operators will be required to handle information and assistance traffic, person-to-person calls, collect calls, and other classes of calls which do not lend themselves to customer handling, as well as such calls as the customers may not wish to dial themselves.

Plant Components Designed for Service at Minimum Cost

It will, of course, be evident that telephone service can be furnished at minimum cost only if the components from which the telephone plant is constructed are designed so that each will perform its function properly at the lowest possible cost. This rarely means minimum first cost, instead, it means minimum annual cost considered over a long period of years, and taking account of initial cost, length of life, and costs of operation and maintenance. This is, therefore, the goal of telephone plant designers and manufacturers.

Desire to progress toward this goal was one of the major reasons which led the founders of the Bell System to establish the Western Electric Company as the principal supplier of apparatus for use in the Bell Telephone System. Today, designs of the articles which the operating telephone companies want and which Western Electric Company manufactures are produced in Bell Telephone Laboratories in close co-operation with both user and manufacturer. The basic objective for each article manufactured is to meet service requirements at minimum annual cost.

In telephone research today, considerable attention is being given to the study of materials, both as regards the fine details involved in their structure and their over-all physical properties. The characteristics of old materials and methods of improving them are thus studied. Also studied are the synthetic new materials now being produced in profusion by many manufacturers. The designers of telephone apparatus, used both indoors and outdoors, work very closely with researchers, manufacturers, and users. In past years, this combination has been a potent factor in holding down annual costs and can be counted on for equal potency in the future. Some examples are as follows:

Telephone Instruments. Modern telephone instruments are designed for long life expectancy and a very low rate of incidence of troubles. Their initial cost is also low. To obtain a design from which long life could be expected required much attention to the choice of materials and many tests of these materials calculated to disclose deteriorations to be expected from aging and environmental conditions. To keep the rate of trouble expectancy low, the designs were subjected to severe mechanical tests as well as exposure to extremes of humidity and temperature. In the factory testing of the Bell System telephone set of latest design, the acoustical properties are exhibited in graphic form on a television-type screen so that the performance of each set is compared directly with the standardized requirements. Thus a simple control mechanism is attained which insures high quality and uniformity.

Vacuum Tubes. Early vacuum tubes had short life and were prone to fail fortuitously. Research and improved design later produced vacuum tubes which could be relied on to give good service for ten years or more with very little danger of failing while in service. These improvements saved millions of dollars in tube replacements. The improved reliability has made it possible to develop systems for long-distance transmission in which performance depends on the unfailing operation of thousands of vacuum tubes. These tube developments finally led to the spectacular achievement of sufficiently long-lived tubes to permit sealing them inside of submarine cable repeaters with an expectancy that cable repairs due to repeater failures would be less frequent than once in 20 years.

Telephone Poles. Before installation, telephone poles are today impregnated with chemicals which prevent fungus growth and rotting. This has increased the life of poles as much as tenfold. Impregnation of poles proves in handsomely, even though the process about doubles the cost of a pole.

Telephone Drop Wire. The insulated conductor drop wires used for connections to the premises of telephone customers have been continuously improved over the years. Largely since the last war, a new design has been introduced which costs very little more but gives good service for more than twice as many years.

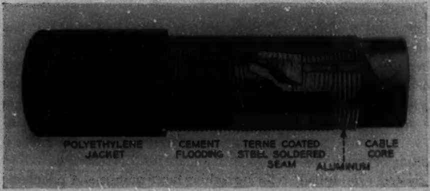
Plastic Sheathed Telephone Cables. This sheath replaces the lead formerly employed. A preferred form of the new sheath, called "Stalpeth," consists of a thin layer of aluminum plus another thin layer of steel with a soldered seam giving a tight enclosure, over which a layer of polyethylene is extruded. This new sheath, as shown in Figure 8, eliminates dependence on the supply of lead and offers promise of worth-while economies.

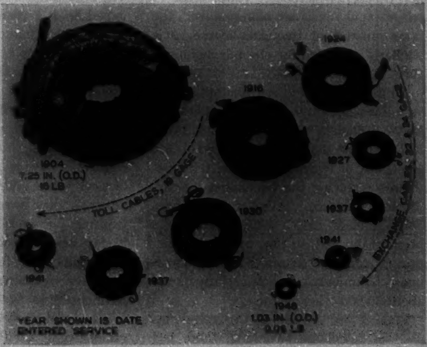
Telephone Relays. When a telephone customer dials a call, often it is required that a thousand relays must operate correctly to make the connection to the desired telephone. In past years, many improvements have been made both in the physical designs and in the circuitry of telephone relays. A new design of telephone relay with wire springs, which has been under development since the last war, has just been announced in which lower first cost, much longer life, and a very much lower trouble expectancy have been simultaneously achieved.

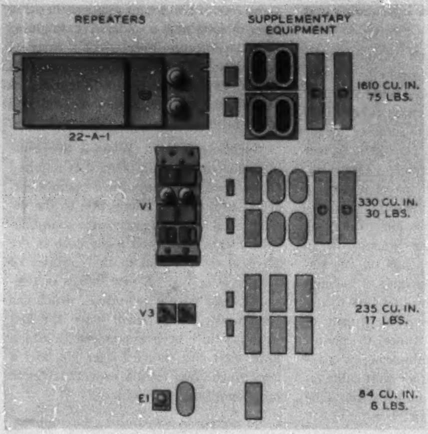
Transmission Apparatus. Great progress has been made in reducing the cost of transmission apparatus, both components and assemblies. These cost reductions have resulted from the utilization of improved materials, superior methods of design, and careful tailoring of the designs to the jobs which each piece of apparatus must perform. The cost reductions usually have been accompanied by considerable reductions in size. As examples, Figure 9 shows the evolution which has taken place in loading coils, cost reductions being comparable to the size reductions. Figure 10 illustrates a similar evolution for voice-frequency telephone repeaters. The El repeater shown at the bottom is a negative impedance device designed particularly for use in local telephone plants where it is now finding wide application. While well suited for this application, its performance capabilities are somewhat limited as compared with

Figure 8 (below). Cut-away view of Stalpeth cable showing different types of protective material

Figure 9 (right). Reduction in size of loading coils, 1904–1952







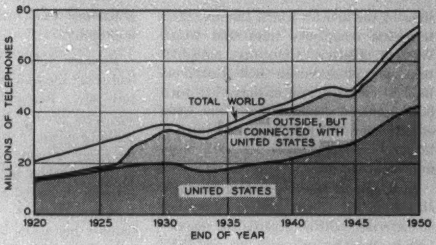


Figure 10 (left). Two-wire repeater development

Figure 11 (above). Growth in telephones

the other repeaters shown, all of which are of the 22-type.

Some Services Associated with Telephony

The recognition that the plant required for telephone service could also effectively give other closely related services led, in the very early days of the telephone, to the superposition of d-c telegraph channels on the telephone wires. These telegraph channels were leased to private customers and to the companies furnishing message telegraph service, and were also very useful in maintenance work on long-distance telephone lines.

With the successful introduction of carrier transmission methods in the early 1920's, it became practicable to furnish large quantities of telegraph channels by carrier methods. Today some 18 telegraph channels are obtained from the same amount of frequency space as is regularly provided for one telephone circuit.

To meet a demand growing out of private-line telegraph service, teletypewriter exchange service was inaugurated in December 1931. This service is quite analogous to telephone service, teletypewriters replacing the telephones on customers' premises. A network of telegraph circuits, nation-wide in extent, has been established, so that today any teleprinter exchange customer can, upon request, be quickly connected to any one of about 30,000 other customers.

Making use of some of the techniques used in telephone switching, there has been developed, largely since the war, an improved form of private-line teletypewriter service. The various establishments of a customer are connected by a network of telegraph circuits. Automatic means for telegraph message handling are provided so that telegrams are automatically delivered from any one of the customer's locations to one or more other locations. This is accomplished by prefacing the telegraph messages with simple code addresses.

About 1920 began the development of public address systems. These included a high-quality microphone, a vacuum tube amplifier, and a loudspeaker so that music and the voices of public speakers could be delivered to large audiences. An impressive demonstration of the possibilities was made on Armistice Day, in 1921, when large audiences at Arlington, Va., New York, and San Francisco, Calif., joined in the ceremonies attending the burial of the Unknown Soldier at the National Cemetery at Arlington. The connections between these points were made by specially arranged telephone lines. Subsequently, public address service was furnished on a regular commercial basis. In later years, however, the telephone companies have largely left the furnishing of local public address systems to others, although they have continued to furnish transmission lines wherever demanded. Today wired music, delivered over telephone lines, is furnished regularly to numerous restaurants and other establishments.

When radio broadcasting of sound programs began, the Bell System established *WEAF*, one of the first broadcasting stations. This was sold after the practicability of radio broadcasting for hire had been demonstrated. Then the Bell System confined its efforts to the furnishing of transmission channels similar in characteristics to those for use with public address systems. Today large numbers of broadcasting stations and studios are connected by such circuits into regularly established chain networks.

The Bell System was a leader in establishing picture transmission service. It entered into the development of machines, however, only to the extent of providing machines for exploring the possibilities of the service and for equipping a nation-wide network for the use of the Associated Press, which has remained in regular service since its establishment in January 1935. Another extensive wire network for facsimile service was established early in 1947 for the use of the Air Force in transmitting weather maps. Also, considerable use is made of regular telephone circuits for temporary use in transmitting pictures from newsworthy locations.

The Bell System pioneered in the establishment of the television industry. In 1927, it set up the first public demonstration of intercity television transmission. President Herbert Hoover in Washington was seen and heard in New York by Walter S. Gifford, then president of the American Telephone and Telegraph Company, and an audience of some 50 scientists and journalists. These television images were carried over an open-wire line. The images were of sufficient clarity to permit easy recognition of the subjects.

Subsequently, the Bell System has largely confined its efforts to the development of broad-band channels so that facilities for both local and long-distance television transmission would be available when demanded by the television industry. The coaxial cable system was a logical candidate for this service as well as for furnishing large numbers of telephone circuits. In 1940, the Republican National Convention in Philadelphia was televised and transmitted over a coaxial system for New York broadcast. Regular network television transmission started in 1946 between Washington, D.C., and New York, with Baltimore and Philadelphia joining the coaxial network the following year. In 1947, radio relay came into the picture and this new system was used to connect Boston to the Washington New York network. Since that time there has been rapid expansion of television networks, high-lighted by the opening in August of 1951 of a nation-wide radio relay system and the first coast-to-coast television broadcast.

Radio

The telephone gap between North America and Europe was bridged by radio in 1927, when the first circuit was established between New York and London using frequencies in the long-wave range. Meanwhile, it was found that higher frequencies, in what was called the short-wave band, would on the average give much better service and would be more economical. In 1928 the first circuit in this band was put in service. Bandwidth is available for a great many more circuits in this range and today the low frequencies are used only in case of failure of the high-frequency circuits. Utilizing the short-wave band, there has been a considerable further growth in overseas radio, so that today nearly a thousand calls are handled each day between North America and Europe. Transoceanic radiotelephone circuits now reach practically all the countries of the world, so that telephones in the United States can now be connected with approximately 96 per cent of all of the world's telephones.

Radio service also has been extended to many seagoing vessels. This service started with the steamship *Leviathan* in 1929 and is now regularly furnished to most large ships. Starting in 1934, many thousands of smaller ships in the coastal waters of the United States and in the Great Lakes, as well as several hundred private airplanes, have received telephone service from shore stations by means of radio. All of these services use frequencies in the short-wave band. After the last war, service was added to ships in harbors and rivers, using frequencies in the 150-megacycle range, the same band used for service to motor vehicles in cities.

Commercial radiotelephone service to automobiles in cities and on highways began after the war and its use is rapidly expanding. Today, telephone connections can be made between any telephone in the United States and more than 12,000 mobile stations of which about 20 are on railroad trains. Several hundred thousand other vehicles are equipped with radiotelephone sets for various private services, such as police, taxicab, and so forth.

In this country some use also has been made of radio to provide service during emergencies. About 100 special equipments stored at strategic points, as well as the mobile radiotelephones in telephone company vehicles, serve this purpose. A few radio systems are in service for bridging short water gaps and crossing other difficult terrain.

As already noted in the section on transmission, microwave repeatered radio systems have now been developed to the point where they have become suitable for providing channels adequate for hundreds of telephone circuits or for television service. A microwave system now spans the continent, and branches extend into many sections of the country.

Conclusion

The telephone business represents the fruition of an invention. The growth of the business, in size, quality, and comprehensiveness of service, has been dependent on the surmounting of many technical obstacles. With the surmounting of these technical obstacles, the telephone has played an indispensable role in the development of American business and industry as well as in the development of social contacts. It has also played a vital and distinctive role in the defense of this country in times of war. There has been a great growth in telephony both in the United States and in the rest of the world as shown in Figures 11 and 12. What remains for the future?

From an international point of view, there is need for more ample and technically better telephone service between nations separated by water barriers, and especially between North America and Europe. Between these continents, radio now constitutes the sole means for telephonic communication. Although the service has been considerably improved since it was started in 1927, the quality of service is sometimes unsatisfactory, the performance of short-wave radio channels being greatly dependent on atmospheric disturbances.

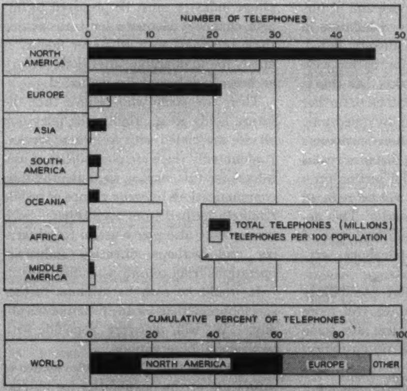


Figure 12 (left). Telephones by continental areas, January 1, 1951

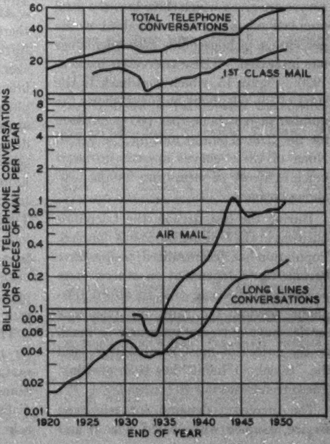


Figure 13 (right). Telephone versus postal service in the United States

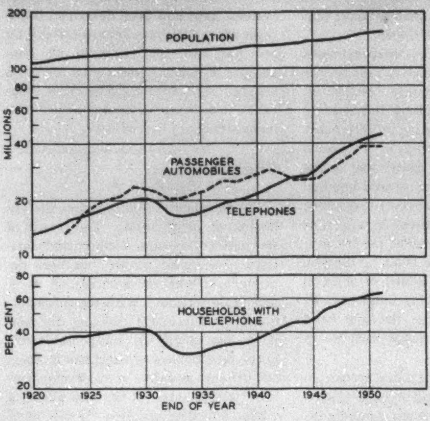


Figure 14. Some statistics of growth in the United States

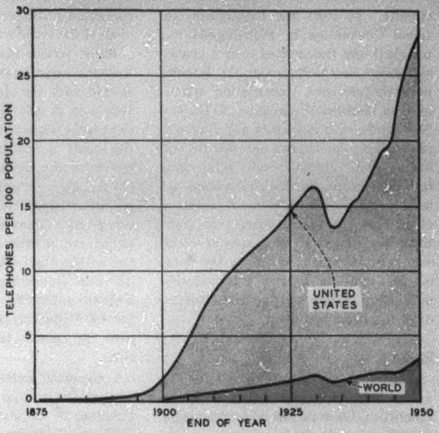


Figure 15. Telephones per 100 population

There are many opportunities for further progress in telephony in the United States. Before considering them, however, let us examine a few statistics. Figure 13 shows growth which has taken place over the years, both in the total number of telephone conversations and in the number of long-lines conversations in one year. For comparison, figures are shown for the number of pieces of first class mail handled by regular service and by air mail service. Figure 14 shows how the number of telephones in the United States has increased as compared with the number of passenger automobiles. Other curves are given which show how the total population has increased and also how the percentage of households with telephones has increased. These curves go back to 1920. Figure 15, which goes back to the start of the telephone business, shows the number of telephones per 100 population for the United States and for the world. None of these curves appears to show much evidence of saturation for the telephone business.

During the last decade, the telephone business in the United States has doubled. Population has decentralized, people have become more dependent on others, more people have become able to afford telephone service, the rates for telephone service have not gone up nearly as much as the prices of most commodities, and the "telephone habit" has been growing. These trends bid fair to continue.

The increased demand for telephone service, coupled with World War II and its aftereffects, has resulted in serious shortages in telephone plant. In the United States it is now estimated that almost 1,000,000 potential customers are waiting for telephones, while about 2,000,000 more would like to have their service upgraded. This is in spite of the fact that, in each year since the war, more than 2,000,000 new telephones, with new outside plant and central office equipment required to serve them, have been added. Telephone service has also been affected. While telephone service in the United States generally can be considered good, weak spots exist in many locations. The first job, therefore, is to provide enough plant so that everyone can obtain the kind of telephone service he desires and the service weak spots can be eliminated.

After these matters are taken care of, there will remain many possibilities of further improving the service and of further combating the effects of inflation in increasing its cost. As already noted, plans have been made for nation-wide automatic telephone service. As this is developed, its first effect will be to further speed up service handled by operators. Later it will permit telephone customers to dial their own long-distance calls. Thus, the time required for setting up a long-distance connection will be reduced to a matter of a few seconds. This development will also benefit the cost of providing long-distance telephone service.

There remain many more possibilities for making telephone service still more useful and convenient. Here are some examples. As the range of customer-dialing of calls is further increased, it is believed that customers will become more appreciative of devices which will be faster and more convenient than the present dial. There will undoubtedly be greater demands for telephones which do not require that a handset be held close to the mouth and ear. For customers temporarily away from their telephones, there is need for improved forms of leave-word service or means for automatically switching the connection to another telephone.

For television, much still lies ahead. Color is coming for reception in the home. Theater television is already demanding wider band channels, adequate for pictures comparable in quality with present motion pictures, both in color and in black and white. The telephone industry must be prepared to furnish channels adequate for transmitting these programs with whatever fidelity is finally found desirable, after both artistry and cost have been fully evaluated.

There are some who believe that the future holds great promise for television service associated with telephone service. Undoubtedly there are possibilities of usefulness for this service, such as providing communications between groups of people remotely located from each other. Such service may also prove useful for observing and perhaps directing industrial operations from considerable distances.

It would be rash to attempt further specific predictions for the future of telephony. We will, therefore, conclude with two general observations.

The first is that there is no evidence of saturation of the possibilities for obtaining further benefits from further developments. Since World War II, there has been, in fact, a greater production of new developments than in any equal span of years in the past. The tempo of development has increased along with an increase in the number of worth-while projects for development. This is indicative of continued progress in the future.

The second observation is that the development of telephony has always been closely intertwined with the evolution of our society. In fact, telephony has played an important part in shaping this evolution. The future evolution of telephony will depend to a great extent on the future evolution of society and the needs for service which this may bring. The record of the past can give us confidence that future developments in telephony will be ample to take care of whatever may be society's needs.

References

General and Historical

1. Carrier Current Telephony and Telegraphy, **E. H. Colpitts, O. B. Blackwell.** *AIEE Transactions,* volume 40, 1921, pages 205–300.

2. In One Man's Life; Being Chapters From the Personal and Business Career of Theodore N. Vail (book), **A. B. Paine.** Harper and Brothers, New York, N. Y., 1921.

3. Telephony Including Automatic Switching (book), **A. B. Smith.** Frederick J. Drake and Company, Chicago, Ill., 1924.

4. The Theory of Probability and Some Applications to Engineering Problems, **E. C. Molina.** *AIEE Transactions,* volume 44, 1925, pages 294–99.

5. The Financial History of the American Telephone and Telegraph Company (book), **J. M. Stehman.** Houghton Mifflin Company, Boston, Mass., 1925.

6. Propagation of Electric Currents in Telephone and Telegraph Conductors (book), **J. A. Fleming.** D. Van Nostrand Company, New York, N. Y., fourth edition, 1927.

7. Speech and Hearing (book), **H. Fletcher.** D. Van Nostrand Company, New York, N. Y., 1929.

8. Beginnings of Telephony (book), **F. L. Rhodes.** Harper and Brothers, New York, N. Y., 1929.

9. Engineering Reports of the Joint Subcommittee on Development and Research, *Reports 1–49,* Edison Electric Institute and Bell Telephone System (New York, N. Y.), 1930–43.

10. Telephone Communication System of the United States, **B. Gherardi, F. B. Jewett.** *Bell System Technical Journal* (New York, N. Y.), volume 9, January 1930, pages 1–100.

11. John J. Carty—an Appreciation, **F. L. Rhodes.** Privately printed, New York, N. Y., 1932.

12. Telephone Theory and Practice (book), **K. B. Miller.** McGraw-Hill Book Company, New York, N. Y., volume 3, 1933.

13. Communication—Past and Present, **Bancroft Gherardi.** *Electrical Engineering* (*AIEE Transactions*), volume 53, May 1934, pages 745–52.

14. Technical Developments Underlying the Toll Services of the Bell System. *Bell System Technical Journal Supplement,* July 1936.

15. A New Telephotograph System, **F. W. Reynolds.** *Bell System Technical Journal* (New York, N. Y.), volume 15, October 1936, pages 549–74.

16. The Bell Telephone System (book), **A. W. Page.** Harper and Brothers, New York, N. Y., 1941.

17. The Future of Transoceanic Telephony, **O. E. Buckley.** *Bell System Technical Journal* (New York, N. Y.), volume 21, 1942, pages 1–19. *Journal,* Institution of Electrical Engineers (London, England), volume 89, part I, 1942, pages 454–61.

18. Telegraphy in the Bell System, **J. A. Duncan, R. D. Parker, R. E. Pierce.** *AIEE Transactions,* volume 63, 1944, pages 1032–44.

19. Reasonable Earnings to Insure the Best Service, **L. A. Wilson.** *Bell Telephone Magazine* (New York, N. Y.), volume 24, Autumn 1945, pages 137–57.

20. A Carrier Telephone System for Rural Service, **J. M. Barstow.** *AIEE Transactions,* volume 66, 1947, pages 501–07.

21. Electrical Communication (book), **A. L. Albert.** John Wiley and Sons, Inc., New York, N. Y., third edition, 1950. See chapter references.

22. Principles and Applications of Wave Guide Transmission (book), **G. C. Southworth.** D. Van Nostrand Company, New York, N. Y., 1950, pages 666–75.

23. Seventy-five Years of the Telephone: an Evolution in Technology, **W. H. Martin.** *Bell System Technical Journal* (New York, N. Y.), volume 30, 1951, pages 214–38.

24. Message Toll Telephone Rates and Disparities. *Report,* National Association of Railroad and Utilities Commissioners—Federal Communications Commission Toll Rate Subcommittee, July 1951.

Switching

25. How the Theory of Probability May Be Applied to Telephone Traffic, **M. C. Rorty.** *Western Electrician* (Chicago, Ill.), volume 36, May 6, 1905, page 356.

26. Machine Switching Telephone System for Large Metropolitan Areas, **E. B. Craft, et al.** *Bell System Technical Journal* (New York, N. Y.), volume 2, April 1923, pages 53–89.

27. A General Switching Plan for Telephone Toll Service, **H. S. Osborne.** *Bell System Technical Journal* (New York, N. Y.), volume 9, July 1930, pages 429–47.

28. Delay Probability Formulae, **C. D. Crommelin.** *Journal,* Post Office Electrical Engineers (London, England), volume 26, 1934, pages 266–74.

29. Tandem Operation in the Bell System, **F. M. Bronson.** *Bell System Technical Journal* (New York, N. Y.), volume 15, 1936, pages 380–404.

30. Etude des Delais d'Attente, **C. Palm.** *Ericsson Technics* (Stockholm, Sweden), number 2, 1937, pages 39–56.

31. Crossbar Dial Telephone Switching System, **F. J. Scudder, J. H. Reynolds.** *Bell System Technical Journal* (New York, N. Y.), volume 18, January 1939, pages 76–118.

32. Crossbar Toll Switching System, **L. G. Abraham, A. J. Busch, F. F. Shipley.** *Electrical Engineering* (*AIEE Transactions*), volume 63, June 1944, pages 302–09.

33. Automatic Ticketing of Telephone Calls, **O. A. Friend.** *Electrical Engineering* (*AIEE Transactions*), volume 63, March 1944, pages 81–88.

34. The Life and Works of A. K. Erlang (book), **E. Brockmeyer, et al.** *Transactions,* Danish Academy of Technical Science (Copenhagen, Denmark), number 2, 1948.

35. The "Director" for Automatic Telephone Switching Systems, **Arthur Bessey Smith.** *AIEE Transactions,* volume 67, 1948, pages 611–19.

36. General Aspects of Community Dial Office Equipment, **Albert Burkett.** *AIEE Transactions,* volume 69, part II, 1950, pages 1540–43.

37. Basic Theory Underlying Bell System Facilities Capacity Tables, **A. L. Gracey.** *AIEE Transactions,* volume 69, part I, 1950, pages 238–43.

38. The Number 5 Crossbar Dial Telephone Switching System, **F. A. Korn, James G. Ferguson.** *AIEE Transactions,* volume 69, part I, 1950, pages 244–54.

39. Fundamentals of the Automatic Telephone Message Accounting System, **John Meszar.** *AIEE Transactions,* volume 69, part I, 1950, pages 255–69.

40. A Full Automatic Private-Line Teletypewriter Switching System, **W. M. Bacon, G. A. Locke.** *AIEE Transactions,* volume 70, part I, 1951, pages 471–78.

41. Design of Switching Circuits (book), **W. Keister, A. E. Ritchie, S. H. Washburn.** D. Van Nostrand Company, New York, N. Y., 1951, 556 pages.

42. Translators and Identifiers in Switching Systems, **H. H. Schneckloth.** *Bell System Technical Journal* (New York, N. Y.), volume 30, July 1951, pages 588–625.

43. Automatic Switching for Nation-wide Telephone Service, **A. B. Clark, H. S. Osborne.** *AIEE Transactions,* volume 71, part I, September 1952, pages 245–48.

44. Nation-wide Numbering Plan, **W. H. Nunn.** *AIEE Transactions,* volume 71, part I, September 1952, pages 257–60.

45. Fundamental Plans for Toll Telephone Plant, **J. J. Pilliod.** *AIEE Transactions,* volume 71, part I, September 1952, pages 248–56.

46. Automatic Toll Switching Systems, **F. F. Shipley.** *AIEE Transactions,* volume 71, part I, September 1952, pages 260–69.

Transmission

47. Inductive Disturbances in Telephone Circuits, **J. J. Carty.** *AIEE Transactions,* volume 8, 1891, pages 100–26.

48. Telephone Repeaters, **Bancroft Gherardi, Frank B. Jewett.** *AIEE Transactions,* volume 38, 1919, pages 1287–1345.

49. Telephone Transmission Over Long Cable Circuits, **Alva B. Clark.** *AIEE Transactions,* volume 42, 1923, pages 86–97.

50. The Transmission Unit and Telephone Transmission Reference Systems, **W. H. Martin.** *Bell System Technical Journal* (New York, N. Y.), volume 3, July 1924, pages 400–08.

51. Measurement of Telegraph Transmission, **H. Nyquist, R. B. Shanck, S. I. Cory.** *AIEE Transactions,* volume 46, 1927, pages 367–76.

52. Carrier Systems on Long Distance Telephone Lines, **H. A. Affel, et al.** *Bell System Technical Journal* (New York, N. Y.), volume 7, July 1928, pages 564–629.

53. A Carrier Telephone System for Toll Cables, **C. W. Green, E. I. Green.** *Electrical Engineering* (*AIEE Transactions*), volume 57, May 1938, pages 227–36.

54. A Twelve-Channel Carrier Telephone System for Open-Wire Lines, **B. W. Kendall, H. A. Affel.** *Bell System Technical Journal* (New York, N. Y.), volume 18, January 1939, pages 119–42.

55. The Conquest of Distance by Wire Telephony; a Story of Transmission Development From the Early Days of Loading to the Wide Use of Thermionic Repeaters, **T. Shaw.** *Bell System Technical Journal* (New York, N. Y.), volume 23, October 1944, pages 337–421.

56. An Improved Cable Carrier System, **H. S. Black, F. A. Brooks, A. J. Wier, I. G. Wilson.** *AIEE Transactions,* volume 66, 1947, pages 741–46.

57. General Mobile Telephone System, **H. I. Romnes, R. R. O'Connor.** *AIEE Transactions,* volume 66, 1947, pages 1658–66.

58. Progress in Coaxial Telephone and Television Systems, **L. G. Abraham.** *AIEE Transactions,* volume 67, 1948, pages 1520–27.

59. Technical Aspects of Experimental Public Telephone Service on Railroad Trains, **N. Monk, S. B. Wright.** *Proceedings,* Institute of Radio Engineers (New York, N. Y.), volume 36, September 1948, pages 1146–52.

60. Local Wire Video Television Networks, **C. N. Nebel.** *AIEE Transactions,* volume 69, part II, 1950, pages 1451–60.

61. The N-1 Carrier Telephone System,

R. S. Caruthers, et al. *Bell System Technical Journal* (New York, N. Y.), volume 30, January 1951, pages 1–32; volume 30, April 1951, pages 418–46.

62. A SUBMARINE TELEPHONE CABLE WITH SUBMERGED REPEATERS, J. J. Gilbert. *Bell System Technical Journal* (New York, N. Y.), volume 30, January 1951, pages 65–87.

63. A BROAD-BAND TRANSCONTINENTAL RADIO RELAY SYSTEM, T. J. Grieser, A. C. Peterson. *Electrical Engineering*, volume 70, September 1951, pages 810–15.

64. EVOLUTION OF INDUCTIVE LOADING FOR BELL SYSTEM TELEPHONE FACILITIES, T. Shaw. *Bell System Technical Journal* (New York, N. Y.), volume 30, 1951, January, pages 149–204; April, pages 447–72; July, pages 721–64; October, pages 1221–43.

No Discussion

Historic Highlights in Developing the Radio Broadcasting and Television Arts

O. B. HANSON
NONMEMBER AIEE

Synopsis: This is a progress report and review of the outstanding pioneering achievements behind the development of radio broadcasting and television as effective services to the public and the nation. Growth of both of these arts receives separate attention in chronological order; discussion also is included of their inherent relationship. Coverage is given not only to the principal phases of development but also to the possibilities of the future.

ALMOST from the beginning of creation, man has sought to extend the power of his senses and to enlarge his capacity to perceive and respond to the world around him. Until a few centuries ago, these instinctive strivings could utilize only the limited powers of the normal human senses and bodily capacities, unaided by scientific devices, so that adventurers wandered over the face of the earth on difficult journeys, and sailed the seas in lonely ships to learn more of the nature of the world and to come into contact with the inhabitants of far-off places.

First utilized by men to extend their natural power of communication with one another were writing and printing. The less gifted and less adventurous spirits could then appreciate the experiences and philosophy of those who were more fortunate in their mental or physical capacities, and thus could satisfy their own instinctive desires for wider participation in the world's affairs.

The Scientific Technique

It was little more than a century ago—a tiny fraction of time compared with the ages that had gone before—that man became adept in applying the scientific technique. Inventive minds began to use scientific laws and principles for constructing devices which would fulfill many ancient human yearnings. Machines and other apparatus were made to multiply many-fold the capabilities of human hands and muscles; optical instruments began to add to the power of vision; railroads, steamships, and automobiles multiplied the powers of locomotion and provided the means of satisfying more readily the age-old desires for visiting other lands and places; electrical communications devices brought distant friends and relatives closer together.

Shackles that chained man to the limited sphere of his own mind and his immediate neighborhood dropped one by one. Today he can move rapidly, easily, and at will; he can increase by thousands of times the power of his hands and arms; he can send his voice by radio to other men around the world, and hear them in turn.

The last shackel now has been broken. Through television, man's eyesight promises to become all-embracing and worldwide. Television provides a new force to the world. What the power to extend vision will mean ultimately in the stream of human life is hard to say. Its progress has been phenomenal, but the eventual extent of its influence can no more be foreseen than were the vast social effects of electricity inherent in the voltaic cells of the early physicists or in the experiments of Faraday.

Challenging the Imagination

The most audacious imagination could not have envisaged the many ramifications of electronic devices to have grown out of Edison's first observation of electron emission and the first transmissions by Marconi of wireless signals.

Theories, facts, and inventions on which radio and television are based go back a long way into the past. For the science of broadcasting is founded on the science of electricity, and thus its beginnings spark from Franklin's kite and its history follows the lines laid by Volta, Faraday, Edison, Steinmetz, and scores of other researchers, experimenters, and inventors. This latter group includes Maxwell, Hertz, Lodge, Branly, Fleming, Marconi, DeForest, and Armstrong.

Study of the phenomena of electric waves and their behavior led to experimentation with the induced currents which radio now utilizes to send around the world sound vibrations of music and the human voice. Marconi was the first to put these currents to practical work. But it was James Clerk Maxwell, in 1867 at the University of Edinburgh, who enunciated the original theory of ethereal waves. This theory, however, was in the realm of pure science and made virtually no stir in the world at large. It is doubtful if Edison, engrossed with electrical experiments 8 years later, had ever heard of the Maxwell theory.

Milestones of Progress

Meanwhile, in Germany Professor Heinrich Hertz began testing Maxwell's theory in a series of experiments, and discovered that the ether would, as Maxwell had said, transmit electromagnetic waves. When he shot an electric spark across a gap, certain oscillations occurred in the terminals. These oscillations set up waves in the surrounding ether, affecting any conductor in the magnetic field. He further established the fact that these waves would travel through the air, even if walls and partitions intervened.

The next important advance was made by another professor, a Frenchman named Edouard Branly of Paris. He noticed that a short glass tube filled with metal

---

Paper 52-261, recommended by the AIEE Communication Division Committee and approved by the AIEE Committee on Technical Operations for presentation at the Centennial of Engineering—1952, Chicago, Ill., September 10–12, 1952. Manuscript submitted July 9, 1952; made available for printing August 8, 1952.

O. B. HANSON is with the National Broadcasting Company, New York, N. Y.

filings, ordinarily a poor conductor of electricity, when influenced by Hertzian waves became a good conductor. These Hertzian vibrations, even though very weak, caused the filings to cling together, or cohere, thus lowering its resistance to the flow of electric current. Branly called this tube a "coherer" because of this action, and this device became the first "detector" used by the wireless pioneers.

In 1881, a number of scientists, individually but almost simultaneously, attempted to utilize the Hertzian waves for transmitting electromagnetic signals. In England, Sir William Preece and Sir Oliver Lodge experimented in their laboratories, as did Professor Branly in Paris. Similar endeavors were made by Professor Augusto Righi in Italy.

Success of Marconi

Another Italian experimenter was a lad of 20 whose skill and enthusiasm marked him for success. In 1896, this young man, whose name was Marconi, went to England and took his transmitting apparatus to Salisbury Plain, in Wiltshire. There he succeeded in sending a message without wires over a distance of 2 miles. This was not far, as compared with the terrific leaps of Hertzian waves in the years to come, but it was far enough to astound the believers and confound the unbelievers.

From that date, the possibilities in radio vibrations were rapidly developed. In 1897, Marconi was sending his signals 24 miles; in 1898, a yacht race was reported by wireless to stations 30 miles away; in 1899, Marconi flashed signals across the English Channel. Two years later the inventor sent the first signals across the Atlantic.

Henceforth, the science of radio was to follow a sporadic course of progress along a route studded with inventions. Fessenden's electrolytic detector replaced the coherer, and Marconi outmoded Fessenden's device with his own magnetic detector. Braun and others followed with crystal detectors that outmoded earlier models. Then, in 1904, Fleming of England devised a valve detector; and in 1906, DeForest came along with his 3-element electron tube, an invention nearly as epochal as Marconi's original apparatus.

The immediate effect of DeForest's tube was to revolutionize radio transmission and reception. It also improved telephonic communication. The Bell Telephone Company found in this tube a powerful booster for long-distance lines, eliminating many clumsy and costly relays, and extending the range of long-distance calls from 1,500 to 3,500 miles.

During this period, radiotelephony also began showing promise, but like the radiotelegraph, its equipment was large, complicated, and unwieldy. Fessenden achieved considerable success at a station in Brant Rock, Mass., and this inspired other experimenters in America and Europe to attempt broadcasts of the human voice.

Radio Broadcasting Takes Shape

Technically, there is no clear line of demarcation beyond which the early development of radiotelephony branched away from point-to-point applications and headed toward the new field of a mass communication service. Facilities for broadcasting were on the road to development long before the initiation of a broadcasting service to the home.

The idea that eventually was to set radio broadcasting on its forward march was conceived in 1916 by David Sarnoff, then associated with the Marconi Wireless Telegraphy Company of America and later destined to become president and then chairman of the board of the Radio Corporation of America (RCA). His idea called for the manufacture of a "radio music box," which he described in a memorandum to his superiors as having the possibility of making radio a household utility comparable to the piano and the phonograph. The aim was to bring music and the human voice into the home by wireless from stations within range. It took time, but virtually all of young Sarnoff's suggestions eventually became facts.

Broadcasting in the sense that we know it today, as a regular service for disseminating news and entertainment, began officially on November 2, 1920. On the evening of that day, station *KDKA* in East Pittsburgh, which until then had been operated experimentally by Dr. Frank Conrad of the Westinghouse Electric and Manufacturing Company, commenced the first regular broadcasting service by putting on the air the returns of the Harding-Cox presidential election.

This event, which allowed amateur radio enthusiasts to learn of Harding's election the moment it was announced, dramatized the possibilities of broadcasting as a medium of instant mass communication and brought it within the full focus of public attention.

Broadcasting received further stimulus a few months later on July 4, 1921, when RCA broadcast a running account of the Dempsey-Carpentier world's heavyweight championship fight, an event of international interest. On that occasion a listening audience was organized through the National Amateur Wireless Association, whose membership outfitted more than 100 halls and theaters with receiving equipment.

This broadcast was heard as far away as Florida. A temporary station *WDY* located at Hoboken, N. J. for this first service was subsequently moved to Aldene, N. J., to which point such Broadway entertainers as Eddie Cantor frequently journeyed to make radio performances. These activities stirred the public imagination and led broadcasting pioneers to increase their efforts to extend and improve service.

Mention should be made here of the active interest of the Westinghouse Company in the possibilities of radio broadcasting, for in addition to *KDKA*, it obtained licenses for *WJZ* in Newark and *WBZ* in Springfield in October 1921, with *KYW* in Chicago to follow in the winter of 1922.

On February 24, 1922, RCA closed *WDY*, after entering into an arrangement with Westinghouse to underwrite one half the cost of operating *WJZ*, and in the spring of 1923 acquired complete ownership of the latter station.

Problems were many in those days and, judging by today's standards, broadcasting equipment was crude and far from dependable. A typical operation of the time was that of station *WJZ* at Newark. Its studio was tucked away in a women's cloakroom of the factory. There, behind a sliding curtain, were installed a phonograph, a piano, a chair, a table, and the microphone. Records were broadcast by placing the microphone in front of the phonograph horn.

In those days, electrical phonograph pickups with direct connection to amplifiers were unheard of and the combination of acoustical recording, mechanical pickup, mica diaphragms, tin horns, and carbon microphones produced a variety of distortions that would drive today's listener to distraction.

The truth was that people could get better and quicker news from their papers and much clearer music from their own phonographs, but the magic of hearing actual voices, even blurred and sputtering, and actual music coming over the air, was so novel and strange that radio set owners clamped headphones to their ears for hours.

But better microphones were in the making. One of the first of the improved type was known to broadcasting engineers as the Westinghouse "tomato can," which was what it resembled. It was better than its predecessors, but it was tricky and capricious. On some days it was excellent, on others, particularly rainy days, it produced a whole series of weird noises, hisses, and growls. Fortunately, however, it opened the way before long for improved pickup devices, including the condenser microphone.

Birth of the American System

The basic character of the American system of broadcasting gradually took shape during the first half of the 1920's. On April 19, 1922, *W2XY*, a 500-watt experimental station of the American Telephone and Telegraph Company (AT-&T) went on the air for the first time. The transmitter was atop the 24-story Walker-Lispenard building in New York City. The program consisted of the reading of extracts from the daily papers. On July 26, a limited commercial license (*WBAY*) was issued for operation on 360 meters and with it began regular program transmission. These call letters were changed to *WEAF*, and later to *WNBC*.

A single studio was provided, adjacent to the transmitter, of regular office-type construction minus sound proofing or acoustical treatment other than wall draperies and a carpeted floor. Double button carbon microphones were used for program pickup purposes.

Operating with a single studio proved to be so great a handicap that two new studios were constructed on the fourth floor of the telephone building at 195 Broadway. These were built with the adjacent walls separated by an air space of 6 feet to provide some measure of sound proofing. felt was used on the ceiling, draperies on the side walls, and on the floors were spread removable rugs for acoustical control. An announcer's booth was located in a cubicle between the studios and double plate-glass windows permitted the announcer to follow the action in the studios and to make his announcements in a relatively quiet enclosure.

These pioneer broadcasts generated so much interest on the part of the public that AT&T and The Western Electric Company were deluged with orders in the New York area alone for the purchase of some sixty transmitters from potential broadcasters who perceived the latent power of this new medium for publicizing ideas and selling merchandise. Even though the Department of Commerce, the licensing agency of those days, recognized the need for and made available a second wave length, 400 meters, for broadcasting, the telephone company decided that rather than be a party to fostering a situation which would lead to eventual chaos (with scores of stations attempting to operate by sharing time on two wave lengths) a much more desirable and practical arrangement would be to set up and operate a toll broadcast service, available for use by any legitimate subscriber who wanted to project a message to the public. This was the beginning of our American system of commercial broadcasting.

The first commercial program over *WEAF* was broadcast on August 28, 1922, under the sponsorship of the Queensboro Corporation, a real estate organization in Jackson Heights, Long Island, N. Y. The charge was $100 for 10 minutes. This commercial program was tremendously important because it led to large-scale advertising on radio, and thus tapped a rich mine of revenue by which broadcasting could be financed. It meant that radio could be delivered without cost to listeners.

This American method of financing broadcasting, with private industry and business as customers for radio time, has eliminated the economic threat of government ownership and has made possible the development of broadcasting without political control and without restrictions save those of prudence and good taste.

Formation of *NBC* Network

Station *WEAF*, after taking this epic lead, continued as one of the foremost pioneers in American broadcasting. It was my good fortune to have been a member of its initial engineering staff and to have participated in many of the *WEAF* "firsts" that made broadcasting history. *WEAF*, which today operates under the call letters of *WNBC*, made the first pickup from a theater, a broadcast from the Capitol Theater on November 19, 1922; the first Presidential address by radio when President Harding spoke on June 21, 1923; the first broadcast from the White House on February 22, 1924, with a talk by President Coolidge; the first political convention broadcast in June, 1924, when the Republicans met in Cleveland to nominate Calvin Coolidge. That was followed by the Democratic Convention in old Madison Square Garden, remembered for the Alabama delegation's dogged "24 Votes for Underwood."

The popularity of these early broadcasts from New York soon brought about a demand from listeners in other areas to hear these programs via stations in their own cities. On January 4, 1923, *WEAF* arranged to feed a special program to *WNAC* in Boston, Mass., for simultaneous broadcast by both stations. This was the first attempt at network broadcasting. Toward the end of the year, a regular network was in operation with *WEAF* as the key station feeding daily programs to *WJAR*, Providence, R. I., and *WCAP*, Washington, D. C. The following year saw five additional stations added and the network grew at an accelerated pace with each succeeding year, so that at the time of the formation of the National Broadcasting Company (NBC) on November 1, 1926, the NBC Network consisted of nineteen stations. The best available statistics at that time indicated that 5,000,000 homes were radio equipped. The day had passed when the radio receiving set was a novel plaything; it had become an instrument of service, a household utility. There could be no doubt that the future of radio broadcasting would be dependent largely upon the character of the programs and the extent of coverage.

It soon became apparent that coverage was insufficient to satisfy the demands of the growing radio audience. Station owners, particularly in cities where there was keen competition, pressed for network affiliations. An early RCA network with *WJZ*, New York, as the key station, was placed under NBC operation. It became known as the "Blue" network and the original NBC group as the "Red" network. Incidentally, the Red and Blue designations grew out of the colors of the crayons used by the telephone company to portray the network lines on their charts. In 1941, the Blue network was to be sold to the American Broadcasting Company. Meanwhile, in 1927, the Columbia Broadcasting System had formed a competitive network. Programs from the major business and entertainment centers of the nation were thus made available for simultaneous listening by audiences in additional areas by three major networks.

The rapid growth of network broadcasting made it mandatory for NBC to expand studio and control facilities. Space was leased at 711 Fifth Avenue, close to the theatrical and musical centers of activity, and a new studio plant was built consisting of eight studios with soundproofing, air conditioning, and a master control room. Each studio was provided with individual booth and monitoring facilities. This installation went on the air for the first time on October 1, 1927. The design of this multiple-studio plant did much to establish methods of sound isolation and improved studio acoustics.

With the expansion of the commercial variety-type show, calling for many hours of rehearsal for each hour broadcast, it soon became evident that the new studio plant would become inadequate in the not too distant future. Consideration of this led to some long-range planning and the eventual construction of the elaborate NBC Radio City studios. These incorporated the last word in architectural design, sound isolation, acoustic treatment, air conditioning, and technical facilities. When the studios were opened on November 11, 1933, they were capable of accommodating 48 separate program originating points and 14 outgoing channels through an automatic pre-set switching system. This plant is still the main headquarters of NBC.

Networks Span Nation

Gradually, the radio networks were extended from East to West and by December 28, 1928, NBC conducted the first coast-to-coast broadcast over permanent network lines, although a one-time broadcast from the west coast had been made in 1924.

At this time, there were approximately 700 broadcasting stations in the United States, and the expanding laboratories of the radio industry, aware of the tremendous possibilities present, worked virtually around the clock improving the quality of broadcasting equipment. This meant better microphones, acoustical treatment and sound insulation, giant water-cooled electron tubes, and high-powered transmitters of 50,000 watts output in place of the 500 watts with which broadcasting began.

Broadcast Receivers

Radio receiver development kept pace with improved transmitting facilities, and with the perfection of the heater-type electron tube, storage space and bulky and expensive "B" batteries were eliminated and the "all-electric" receiver came into universal use. Multiple-tuning controls were combined into one and improvements in circuit sensitivity marked the demise of the long outdoor aerial and the birth of the "built-in" antenna.

It may be recalled here that Westinghouse inaugurated its broadcasting service at *KDKA* in 1920 and in other cities in 1921 to create a market for radio apparatus. Commercial broadcasting as we now know it was not contemplated. Receivers offered for sale in that period were relatively crude, consisting of crystal sets with no electron tubes, or one-tube sets, some two-tube sets, and an expensive set utilizing four tubes and a horn-type loudspeaker.

Within the few years that followed, the efficiency of many new models of receivers was improved by the utilization of regeneration, but in 1924 the first really great improvement in radio receivers took place with the introduction of the superheterodyne receiver. It first had been used by Major E. H. Armstrong when he was an Army officer engaged in improved methods of detection of enemy radio signals during World War I. When this circuit was adapted to broadcast receivers in 1924, a change was launched which led to the almost universal use of this type of circuit for public reception of broadcasting.

About this period, other noteworthy developments in receivers took place, including the superregenerative circuit, circuits utilizing reflex principles, and others utilizing feedback neutralization principles, the latter identified with Dr. Alan Hazeltine. It was called the neutrodyne. The superheterodyne circuit reduced the frequency of an incoming signal so that it could be more efficiently amplified, and gave superior control over performance characteristics.

All of the circuits discussed were used in broadcast reception but through a period of years the superheterodyne superiority and advantages created wide public acceptance and the almost complete elimination of other types of circuits.

The next outstanding improvements in receivers were the introduction in 1927 and 1928 of automatic-gain control and a-c operated tubes which improved and simplified receivers.

Present-day receivers, representing steady improvements through the years, differ little in basic principles from the sets of 20 years ago, when the important basic improvements had been introduced.

The modern receiver, incorporating simple controls to select the desired station and adjust the volume and tone, is a remarkably efficient, inexpensive, and reliable instrument constituting a monument to American ingenuity and the system of free competitive enterprise.

Improved Programming

Broadcasters held public attention with improved programming and added new laurels to the art of outstanding coverage of news and special events of national interest. I recall an appropriate saying coined during that period by David Sarnoff of RCA: "The richest man cannot buy for himself what the poorest man gets free by radio." This was indeed true, and still applies to American radio as well as to television. Radio brought to its millions of listeners the first broadcasts directly from the stage of the Metropolitan Opera, the first transmitted symphony to be conducted by Maestro Arturo Toscanini, and the first broadcast performances of the great and the near-great of the stage, screen, and musical world. All of the broadcasts called for special engineering and devices. Seven hours of broadcast time were devoted to the first complete radio coverage of the inauguration of a President of the United States when Franklin D. Roosevelt took the oath of office on March 4, 1933.

From an engineer's viewpoint, broadcasting truly came of age during the 1930's. Its rapidity of growth was unprecedented in American industry and the quality of its service, together with its power of immediacy for mass communications, made it a vital component of our way of life. By 1941, on the eve of America's entrance into World War II, there were more than 56,000,000 radio sets in this country and nearly 1,000 stations, many of which were affiliated with four national networks. On the world scene, radio broadcasting had telescoped time and space, bringing virtually every land within earshot of the powerful transmitters here and abroad.

Radio's Wartime Role

During the next 4 years radio played a key role in the global conflict. Our government and armed services were afforded direct and simultaneous contact with the people of the United States and other nations. Two decades of development of the American system of broadcasting, based upon American traditions of free speech and free enterprise, had prepared radio for its biggest job. It rallied the nation, broadcast vital information, and helped to unite freedom-loving peoples everywhere in the single purpose of winning the war.

Networks altered their regular schedules to present Presidential addresses and announcements of national interest, cancelling commercial revenue which they would have received. For example, one of the major networks alone devoted more than 10,000 hours of network time to the war effort. An arrangement was made with the government through which the network's high-powered short-wave facilities and personnel were devoted to a daily average of 15 hours in eight languages: French, German, Italian, Swedish, Danish, Spanish, Portuguese, and English. News programs were beamed across two oceans and to Latin America, and included interviews and special events, as well as entertainment. Special programs were beamed to American troops in various parts of the world.

At the war's end, recognition of the potential value of this overseas broadcasting as a tool to spread abroad a true picture of our country, what it stands for, and what its aims are, and in countering the false propaganda of our enemies, resulted in a greatly augmented schedule of broadcasts under the aegis of the State Department and an increase in the power and number of transmitters serving the world.

Members of our engineering staff assisted government agencies on many important wartime projects, most of which were secret in nature. In all, more than 30 different research projects were undertaken for government and military services.

Research and Development

While primarily engaged in program transmissions, radio broadcasters devoted considerable time through the years in research and development to find the answers to problems that were continually arising and to improve the technical quality and reliability of their broadcasts.

The subject of sound recording is one that has been under continuous study from the beginning in order that each new recording might be tried and new methods and techniques investigated. The "Orthacoustic" disk recording system, now standard in the industry, was an outgrowth of this research. In this method of recording the high frequencies are tilted or "pre-emphasized" in recording and a reverse curve employed during play back, resulting in improved quality and very low surface noise. Tape has been subjected to considerable investigation in recent years and today is widely used in our broadcast operations.

NBC's engineers developed a limiting or automatic audiogain control amplifier to prevent overloading transmitters and to permit an average increase in modulation percentage. Manufacture of this unit was undertaken by RCA and today many of the nation's broadcasting stations are equipped with them.

A study involving co-operation with many foreign countries that extended over several years was made in co-operation with RCA Communications of the correlation of transoceanic radio reception, fading and signal strength, with sunspot and magnetic activity. The resulting information proved very useful in connection with overseas pickups, in forecasting reception conditions, and in determining choice of frequencies to be used in any given period. The first shortwave pick-up from Europe over NBC networks took place on February 1, 1929, when the London Symphony Orchestra was broadcast.

In 1933, the first national survey of radio station coverage performance was made on more than 90 stations, and contours established. These surveys were augmented in the years which followed. Much original research was made on the design and functioning of radio towers and antennas to control fading, and this information was made available to the industry generally.

Today's Great Radio Service

While the momentous events which crowded the pages of history during the war years were sharply reflected in the operations and public services of radio, the end of the conflict marked the beginning of a peacetime phase of activities that carried broadcasting to heights undreamed of a generation earlier.

Today, more American communities have a radio station than a daily newspaper. There are, in fact, more than 2,500 radio stations on the air in this country. Of these, nearly 500 operate on the newer FM (frequency modulation) channels.

Frequency Modulation

The first known disclosure of the principle of frequency modulation for the transmission of voice and music by radio appeared in patent applications as early as 1902. Through the years that followed, frequency modulation was a subject of much research and development and considerable application in telegraphy and other nonbroadcast services. During the ensuing years, many patents were issued to many individuals, on frequency and phase modulation. In the 1930's, application of the FM principle to broadcasting was stimulated by numerous investigators and by extensive field tests.

At the beginning of World War II, there were in existence in the United States 110 FM broadcasting stations and about 350,000 receivers. At the conclusion of World War II, the Federal Communications Commission (FCC) moved FM from the 40-megacycle region to the 100-megacycle region, made 100 channels available, and opened the way for FM to grow. Although it grew rapidly during the first postwar years, this growth did not accelerate or continue. To some extent this resulted from the wide public interest in television which at that time was starting its phenomenal growth.

Frequency modulation is used also in the television sound channel with somewhat different standards than for sound broadcasting.

By midsummer of 1952, there were more than 105,000,000 standard and FM radio receivers in the United States. This is nearly double the number in American homes before World War II.

Concurrently with this progress in the sound broadcast field, scores of engineers in many laboratories have been busily engaged in research to effect the eventual union of sight with sound, so that the public may be provided with a completely rounded out medium of communication, entertainment, information, and education. When we note the 109 television stations now in operation and the distribution of some 19,000,000 receivers in the United States, we realize how successful the efforts of these engineers have been. While there are some who believe that the continued growth of television will perforce work to the detriment of radio broadcasting, I am of the firm belief that radio will continue to be of incalculable value to listeners and the nation for a long time to come.

Television

Before 1900, there was no such word as television. Several people have claimed to have coined the word, including a Frenchman named Perski. In preparing material listed in *Annexes Congres Internationale d'Electricité* for August 18–25, 1900, Perski employed the French word for television in its modern meaning. Eleven years later, an unnamed examiner at the British Patent Office, in forming a search file devoted to the subject, used the word in its present anglicized form.

I have intentionally avoided mentioning television's parallel development with radio broadcasting, particularly during the 1930's, to provide a concise view of both radio broadcasting and television progress from their beginnings.

Chronologically, the science and art of television, as of radio, can be dated back a hundred years or more, since both are based upon electrical principles and other fundamental discoveries. Television has tapped virtually all of the reservoirs of knowledge in radio, chemistry, physics, optics, and electronics.

The word television—seeing at a distance—has by common acceptance come to mean the instantaneous transmission and reproduction, either by cable or radio, of moving scenes and pictures. Its operation is far different from that of the telescope, an optical instrument, by means of which a scene is viewed from a distance; from facsimile, which is the electrical transmission of still pictures; and from the simple distribution of moving picture films of a scene to a distant point for reproduction.

Essentially three steps are involved in television: first, the analysis of the light image and its transformation into electrical signals; second, transmission of the electrical signals to points of reception; and third, visual reproduction of the original image from the electrical signals.

Every element, every component, of television has required years of exploration and pioneering. New principles had to be discovered in picture pickup devices, electron optics, electronic amplifiers, broadcasting transmitters, transmitting tubes and antennas, and in methods for synchronizing the transmitted and received pictures. New parts of the radio spectrum had to be explored and harnessed.

Television today rests upon its own solid foundation of research. It encompasses many remarkable scientific advances, each of which in its own right may be regarded as a monument to the progress of science.

The first epic came when Baron Berzelius, a Swedish pharmacist, discovered the element selenium, in 1817. It took 56 years, however, before its importance was established. Then a telegraph operator named Joseph May, employed at a cable station in Ireland, noticed that a resistor made of selenium seemed to conduct current more freely when the sun shone on it. Other investigators found that it was the light and not the heat of the sun that made this change in the conductivity of selenium. May's discovery was to lead, after many years, to the modern photoelectric cell, which converts light waves into electrical signals. The photoelectric cell is a vital component of television.

Other early events of paramount importance in television include the invention, in 1878, by Sir William Crookes of the Crookes' tube and his demonstration of cathode rays; Edison's discovery in 1883 of the "Edison Effect"; and the patenting in 1884 by Paul Nipkow of a rotating scanning disk as the first practical means of analyzing and converting a picture into an electric current suitable for transmission over wires.

Although the fundamental principles of mechanical scanning were revealed before the beginning of the 20th century, practical application for television was impossible before the invention, in 1907, of the 3-element thermionic amplifier by DeForest. During the next 20 years, considerable work was done in the television field utilizing the principle of mechanical scanning, and many improvements were proposed. Successful demonstrations were conducted by John L. Baird in England and Charles Francis Jenkins in the United States, as well as by NBC and Bell Laboratories. These demonstrations were more in the nature of silhouettes, but soon thereafter half-tone pictures were achieved.

The mechanical television system suffered from a number of serious defects. It was optically inefficient; mechanical components rotating at high speeds were involved; and a high intensity light source was extremely difficult, if not impossible, to obtain in practice.

Introduction of Electronics

The final solution to the television problem was found in a different direction, namely, through the use of electronic pickup and reproducing tubes and electron beam scanning.

As early as 1905, Boris Rosing proposed the use of a cathode ray tube as a means of reconstructing television images, and demonstrated by experiment that this was feasible. In 1911, Campbell-Swinton suggested a special cathode-ray tube which would function as an image pickup device. His proposed tube, however, was of theoretical interest only and was never produced.

It was not until the early 1920's that electronic pickup tubes were actually shown to be operative. The first electronic device of this type was the iconoscope, invented by Dr. V. K. Zworykin in 1923. In 1929, Dr. Zworykin, now vice president and technical consultant of RCA Laboratories, disclosed the kinescope picture tube which he had developed for home television sets, and demonstrated the two simultaneously for the first time.

As the forerunner of today's all-electronic television system, this may be regarded as the greatest single advance in television art, for it meant the end of mechanical scanning with its inherent limitations and imperfections.

Another fundamental obstacle in the early days hurdled by research engineers was limited picture definition. This stood in the way of building any kind of television system suitable for satisfactory service. The major cause was the lack of suitable amplifier tubes capable of handling the required frequency bandwidth. The frequency band needed to transmit the television signal of an eminently satisfactory picture is 4,000,000 cycles, or 4 megacycles. Development of amplifier tubes to achieve this is acknowledged to be a brilliant contribution. Without this engineering feat, television as known today would not have been possible.

Related to the development of amplifier tubes was the problem of power-transmitting tubes with the capacity to operate in the very high frequency band suitable for television transmission. By the early 1920's, work had been begun on such tubes and from then until now research has continued and improvements have kept pace with the expansion and needs of the art.

The original iconoscope, invented by Dr. Zworykin, has undergone many changes. It opened the way for the development of the more sensitive orthicon, introduced in 1939, and for the now universally used image orthicon tube, which emerged from wartime research as a revolutionary advance in pickup-tube sensitivity.

The image orthicon, a hundred times more sensitive than earlier camera tubes, solved many of the major difficulties of illumination in television programming and took television out of the studio for 'round-the-clock coverage of all kinds of events. Today's television cameras can now pick up scenes at lower light levels than those required for motion picture film cameras.

Beginning of Field Tests

RCA was the first leading television research organization to conduct extensive tests outside the laboratory. Experimental station *W2XBS* was licensed in 1928 and began operation in that year, with the transmitter located in New York City. Subsequently, NBC engineers took over the operation of this station to continue experiments.

Those were trying and crucial days for the new art. For if sight and sound were to be combined and proceed in the public interest, it was agreed by the pioneers, broadcasting had to keep pace with receiving techniques.

The year 1931 saw the establishment of a landmark in New York, and in television history. Atop the Empire State Building, the world's tallest skyscraper, was erected the transmitting antenna for station

W2XBS as a predecessor to *WNBT*, America's foremost television station. Twenty years later this site was shared with four other leading New York television stations.

As tests continued in the early 1930's engineers succeeded in increasing the quality of transmitter pictures and overcame many of the early problems. Full-fledged field-test television transmissions from *W2XBS*, using the all-electronic television system, were begun on June 29, 1936.

In Service to the Public

Less than 3 years later, on April 30, 1939, came the introduction of television service to the American public. This premiere event was arranged by RCA, in co-operation with NBC, at the opening of the New York World's Fair, appropriately named the "World of Tomorrow." Participating was President Franklin D. Roosevelt as the first Chief Executive ever to be televised. This was to herald the great jumping-off point for television as a service in the United States. There followed events of television coverage that made history.

The war curtailed television broadcasting, but stimulated development of circuits and tubes. When the conflict ended, television's expanding vigor transformed into reality the hopes and dreams of its pioneers.

The extent of television's phenomenal growth may be best brought into perspective by recalling the situation only 6 years ago. In that year, 1946, as the first postwar home television receivers were introduced, there were fewer than 10,000 sets in the entire country. Television broadcasting was conducted on a meager, few-hours-a-week basis and there were six licensed stations.

On January 1, 1947, sets in use by the public numbered 16,476. By January, 1948, the total had reached 189,000. A year later, it was 1,000,000. On January 1, 1950, it was 3,950,000. The 10,000,000 mark was passed in January 1951, and now there are nearly 19,000,000 sets, with an audience estimated at nearly 70,000,000!

From the few television stations on the air 6 years ago the number has risen to 109, covering 64 major market areas. Television-set manufacturers have increased accordingly, from a few pioneers to a total approaching 100.

Today, the radio and television business which provides jobs for more than 500,000 persons in the United States, has become a $5,000,000,000 industry.

This vast growth has occurred despite the so-called television "freeze" which prevented normal construction of new stations from September 1948 until April 1952. The freeze was imposed by the FCC which claimed it was necessary in order to clear up technical problems that had arisen with respect to interference between stations.

Coast-to-Coast Television

One of the most significant events in television history—from the standpoint of the viewing public, broadcasters, and television set manufacturers—was the opening in 1951 of AT&T's new transcontinental radio relay system linking major cities in a coast-to-coast television network.

Previous to this epic achievement, it should be noted, the Bell System had contributed immensely to television by its development of the coaxial cable and radio relay systems to provide for the transmission of television signals between the major cities of the country. These facilities were, and still are, imperative to network operations. The Bell System continues to improve and expand these facilities as new television stations are constructed.

But now that television has gone so far in so short a time, what does the future hold? Does the opportunity exist for extending television service to the hundreds of localities out of range of present coverage? What can the scientists and engineers do about it? What can the industrialists and broadcasters do?

These questions are paramount today whenever television's future is discussed. Fortunately, some of the answers can be found. And as they come before us there emerges a view of vast horizons, so promising that they make the present stage of television appear as little more than the beginning.

An appreciation of the fact that many localities are without cable or radio relay network service, and of necessity will be for a considerable period, prompted the development of kinescope recording to serve such communities. This is a method of photographing, on motion picture film, the television picture as it appears on the face of a special kinescope monitor. This service has been in operation and available to noninterconnected stations since 1947. Further development of the system is being pursued vigorously at this time. The development is concerned with special motion picture camera emulsion characteristics, development technique, and many other facets of motion picture photography and processing, and includes 16 millimeter films, 35 millimeter films in black and white, and in color.

A major expansion in television service can now proceed in the ultrahigh frequency (UHF), in the range from 470 to 890 megacycles, portion of the broadcast spectrum. This official permission to expand, which also lifts the 3-year ban on television station construction, opens the way for television to reach millions of people who have never before known its enjoyment.

Present and prospective television audiences in many communities are destined to discover that the simple combination of letters UHF stands for more channels, more stations, more programs, and far greater television coverage. And (this is important) the addition of UHF can be achieved without diminishing television service on the very high frequencies (VHF), which will continue to reach the present growing television audience.

Foreseen only a few years from now is the day when America will have 50,000,000 television set owners, and there will be stations within range of nearly every community. Estimates by those familiar with the possibilities place the eventual number of television stations in this country at 2,000 with most of them linked in coast-to-coast and regional networks. This will mean television service for practically the entire nation.

Color Television

The question may be asked: "What about television in full color?"

My answer, which can be made without danger of contradiction, is that practically everybody agrees that television in natural color is highly desirable. It has been the goal of our engineers since the early stages of success in developing the all-electronic system of black-and-white television.

At present, an incompatible color television program service is the only system having the approval of the FCC. It is known as a field-sequential method and cannot be received even in black and white on standard black-and-white sets without the added expense of converters and adapters.

RCA-NBC engineers and virtually the entire engineering profession believe that an all-electronic, compatible color system such as that developed by RCA is far superior to the incompatible method. In adopting standards for the incompatible system, the FCC said that it was not discouraging the development of an improved method of color television. Consequently, RCA in co-operation with other radio organizations is continuing to develop the all-electronic compatible method based on mutually agreeable standards, determined by the National Television System Committee. We are more convinced than ever that this method ultimately will attain approval in the public interest.

Some progress was achieved in color television prior to World War II, but it was in the period that followed that eye-opening advances came in the development of the RCA tricolor picture tube and other vital components of a color television system capable of an eminently satisfactory service to the public.

Beginning with a memorable demonstration in October 1946, RCA has shown in one test after another the strides being made toward its goal of an all-electronic, compatible color television system, comparable in excellence with the high standards of black-and-white television and receivable in black and white on existing sets without added devices. Our engineers at NBC have participated in these tests and the public not only has been informed of progress, but also has participated in many of the tests.

Tests conducted in New York and Washington during the summer and fall of 1951 featured RCA's 21-inch tricolor picture tube and showed such improvement that viewers representing the press, the public, and the television industry were moved to acclaim the system.

Thousands of telegrams, letters, and telephone calls were received from television set owners in the New York metropolitan area who viewed the colorcasts in black and white on their home receivers. This was particularly true during the tests in July 1951 and 1952, in response to an invitation to the public to help test the compatibility of the color system by reporting on reception.

Tests of the RCA color television system, employing standards recommended by the television industry, are in progress at present. NBC's studio 3-H in Radio City, New York, equipped with the latest RCA color apparatus, is being devoted exclusively to these tests.

Looking Ahead

Television as we know it today, a service of vast social, economic, and political significance with a potential audience numbering tens of millions, is destined for still greater improvement and expansion. It will become international in scope and before much longer, we hope, compatible color will be added.

The future perfection of the transistor, upon which so many laboratories and research men are concentrating at present, will give an impetus to improvement of both transmitters and receivers, and in fact, to any type of electronic apparatus now using vacuum tubes.

Television, in addition, holds promise of tremendous growth in other fields. Little publicized is the fact that today more than sixty of the leading theaters of this country are equipped to present large-screen television shows and special events. This means that theater television is fast becoming a mass audience entertainment medium. Moreover, RCA is developing a color television projector for theater screens, and this has very good future possibilities, since it is compatible with the black-and-white system.

Development also is going forward on a system of large-screen color television projection, known as Eidophor, which is being introduced in the United States by Twentieth Century-Fox Film Corporation. This system originates from a Swiss invention and has been adapted for color in co-operation with Columbia Broadcasting System.

Another important field is industrial television. It is being developed on closed circuits, as contrasted to broadcast television, and has the basic aim of extending human sight for the benefit of commerce, industry, and education.

Thus, the future of the television art is exceedingly bright. Together with radio broadcasting, television has become a shining symbol of America's vitality and the will of our people to extend science and industry for the greatest possible good, and for the enhancement of our way of life. Engineers everywhere can feel pride in this achievement. It is both a challenge and a bright ray of hope for the future.

No Discussion

The First 50 Years of International Radio Communication

HARADEN PRATT
FELLOW AIEE

ONE hundred and sixty-four years ago Galvani discovered the existence of current electricity, previous knowledge having been confined to static electricity and the spark discharges associated with it. Significant additional fundamental discoveries made by Volta, Oersted, and Faraday provided the knowledge that enabled Morse to develop the practical instrumentality of the wire telegraph which used direct current interrupted in accordance with the code which he devised. Later Bell developed the telephone using the same basic technical information but employing alternating currents having acoustic frequencies.

Faraday discovered and demonstrated that electrical and magnetic phenomena had an influence on light thereby disclosing the existence of a link between these sciences. This invoked further interest in the mechanisms associated with the natural phenomena investigated by Newton, Faraday, and others of their time, which involved the translation of force at a distance, that were not susceptible of much understanding without the conception of the existence of a connecting medium.

Mathematical physicist Clerk Maxwell interested himself in the problems posed by these previous researches and in 1865 published his epoch-making theoretical paper on the electromagnetic theory of electric waves and light.

After studying Maxwell's paper and the findings of earlier pioneers that the spark discharges produced by static electricity

---

Paper 52-263, recommended by the AIEE Communications Divisions Committee and approved by the AIEE Technical Program Committee for presentation at the Centennial of Engineering—1952, Chicago, Ill., September 10–12, 1952. Manuscript submitted June 16, 1952; made available for printing July 28, 1952.

HARADEN PRATT is Telecommunications Advisor to the President, Executive Office of the President, Washington, D. C.

involved the backward and forward oscillations of electric currents, Hertz demonstrated in 1888 that such oscillatory effects in a circuit projected energy through space and induced similar oscillations in another like circuit at a distance. Thus was brought to light a new field in the development of the electrical sciences which later led to the transmission of messages through space without the use of connecting wires.

Few persons, however, thought much about using these new disclosures for practical purposes although the eminent Sir William Crookes predicted in 1892 in considerable detail the use of electric waves in space for telegraphing without wires employing the Morse code. Most investigators concentrated on broadening this field of electric wave motion knowledge through experiments and theoretical analyses. The first outstanding achievement came when Marconi in 1895 sent Morse signals a mile and a half with such waves generated by the spark method employing detecting devices previously invented by others coupled with a contribution of his own consisting of connecting one side of his oscillatory apparatus to the ground and the other side to an elevated conductor which method was also applied at the receiving apparatus. In 1896 he took out the first patent on wireless telegraphy; in 1897 a company was formed to exploit the development; in 1898 wireless communication was applied to naval maneuvers; in 1899 the first wireless distress signals were transmitted resulting in effective rescue at a sea disaster; also in 1899 the international yacht races at New York were reported by wireless for the New York Herald; and in 1901 Marconi sent signals across the Atlantic from England to Newfoundland demonstrating the possibility of using this method for international long-distance communication. These transatlantic tests were interrupted because of a protest by the Anglo-American Telegraph Company which held an exclusive concession in Newfoundland for operating transatlantic telegraph service, and the Marconi Company consequently built its initial radio stations on the American side in Nova Scotia and on Cape Cod. A transatlantic commercial service was eventually opened in 1907 from Clifden, Ireland, to Glace Bay, Cape Breton Island.

The first practical service rendered by the new art, however, was for communication with ships which use became so firmly established that efforts were made at an international conference in 1903 to establish standards to facilitate the handling of correspondence on an international basis. A convention adopting rules and regulations was agreed to by the important maritime countries of the world in 1906 at Berlin. The marine disasters of the S.S. Republic and S.S. Slavonia in 1909 and the S.S. Titanic in 1912, for all of which radiotelegraphy played an important life-saving role, spurred the enactment of legislation to regulate the proper provision of this means of communication for ships and another international radio-telegraph convention was concluded at London in 1912, which provided safeguards by treaty.

The early efforts of Marconi stimulated activity by others and for over a decade and a half inventors came forth with improved devices and systems mostly in Germany and the United States. This period was also accompanied by the promotion of many stock companies to exploit this seeming enticing and glamorous new attraction. Many little radio systems were established, some to compete with the wire telegraph and others to provide communication over difficult routes, such as between the several Sandwich Islands, Catalina Island to the mainland of California, across the outer end of the Gulf of Lower California, within Alaska, and so forth. The latter far outlasted the former and in some of these instances radio still provides service.

It is interesting to note that as these early developments were progressing, a flight of scientific research interest from the field of electromagnetic wave radiation to the newly discovered and captivating fields of radioactivity and electron theory took place. As a consequence, future radio communication development became more the concern of the practical engineer and inventor.

The initial radio system using the spark-type transmitter relied on devices at the receiving terminal which actuated a mechanical telegraph instrument. It was soon found that greater reliability and extended ranges were obtainable with devices actuating a telephone receiver, due to both the greater sensitivity of the latter instrument enabling the easy identification of minute signal currents and to the discriminating ability of the human hearing mechanism enabling desired signals to be identified from interfering effects. The next step was to produce spark trains at the transmitter at a musical tone rate so that the ear of the listener could identify the signals still better and also to take advantage of the phenomenon of acoustic resonance. This development also reduced the high voltages within the transmitting system making easier the application of higher powers. Another trend was to achieve the transmission of a narrow wave band and apply improved selective means at the receiver to enable signals from other stations on adjacent wave bands to be rejected. The spark transmitter was not well adapted for securing this objective, but nevertheless many improvements were made resulting in the German quenched-spark method and the British and American rotating-spark-gap method.

It soon became obvious that a continuous type of high-frequency wave was the answer to many of these problems. Some experimenters were attempting radiotelephony, but as the spark method was entirely unsuitable for that purpose, further interest developed for using continuous-wave transmission methods on this account. The first practical method of generating such currents was by the use of an electric arc in hydrogen invented by Poulsen of Denmark in 1903. It took 7 years before another group of pioneer developers put it to regular use. The second practical method was generation by rotating machines. Many experimenters had built small machines, but Fessenden was the first to cause the building of an alternator in about 1906 that developed a useful amount of power. About 10 years elapsed before the high-frequency alternator was used for regular practical communication service. With these two means successful radiotelephone transmissions were demonstrated prior to 1907.

With the advent of continuous waves, new methods were needed in order to detect them at the receiving terminal. Because of the absence of trains of waves at acoustic frequencies, these currents had to be broken up or modulated at the transmitter or in the receiving apparatus to make them audible in a telephone receiver. Both methods were tried, but manipulation at the receiver became the accepted way. Poulsen used an interrupter with his arc system, the Germans a rotating commutator device called a tone wheel with their alternator, and Fessenden obtained his audio tone by combining his received high-frequency current with another locally generated high-frequency current differing in frequency from the first by the value of the audio-frequency desired. Fessenden called this method the heterodyne and disclosed it in 1907. It stands as one of the most fundamental elements of radio systems, but again some 7 years elapsed before its advantages were fully realized and put into general use.

Thus Marconi's spark generators based on the apparatus of Hertz became outmoded. The Marconi Company attempted to meet the threat of continuous waves by a very ingeneous method called the timed spark, which combined most of the advantages of continuous waves with many of the disadvantages of the spark type. This company made large investments between 1913 and 1917 in transoceanic systems employing both spark and timed-spark transmitters of 300-kw power but they were obsolete before they could be used because successful transoceanic circuits using both arcs and alternators had been previously established by American and German companies.

The course of developments was radically accelerated by the advent of the vacuum tube and the insertion in it by DeForest of the grid electrode in 1906. For many years the device was confined to the detection and amplification of radio signals, but about 1912 it was discovered that the tube could generate high-frequency oscillating currents when suitable circuit arrangements were provided, and Armstrong disclosed how such oscillations could be controlled. This device became popular overnight by providing a very practical local generator which made the heterodyne receiving method with its many advantages a complete success supplanting all other methods for telegraphy. It made possible the momentous demonstration by the American Telephone and Telegraph Company in 1915 of the successful transmission from a high-power vacuum-tube transmitter installed through the courtesy of the United States Navy in its Arlington, radio station of radiotelephone signals to Honolulu and to Paris. This and subsequent developments of continuous wave radiotelephony made possible the establishment of the first commercial radiotelephone circuit between Los Angeles and Catalina Island in 1920. However, the need for development of practical means of generating the high power needed for long-range telephony in the low-frequency portion of the spectrum delayed opening of intercontinental telephone service until 1927 when the first such circuit was established between New York and London. The versatility of the vacuum tube was strikingly revealed by this demonstration as it served for generating, modulating, rectifying, and amplifying alternating currents. The availability of this many purpose device sounded the death knell for the spark, arc, and alternator systems. It also opened up new horizons by providing the means to generate and control frequencies much higher than were possible with arcs and alternators and at useful power levels. While Marconi's first patent described apparatus using waves as short as 10 inches, it was 45 years later before such frequencies became established for regular useful purposes because of waiting for the invention and development of instrumentalities such as the vacuum tube, which were needed to handle them successfully.

The vacuum tube was responsible for another epoch which started soon after 1920 when it was found by radio amateur enthusiasts that waves shorter than 100 meters could be used over long distances. This was possible because they were bent back by ionized layers in the upper atmosphere thereby overcoming the obstacle of the curvature of the earth. It was quickly found that this new transmission method provided advantages over the existing low-frequency method which was plagued with many difficulties. The facilities were very expensive to build and maintain because poor radiation efficiency demanded the use of very high power and many tall masts or towers to support massive antennas. The radiations could not be beamed to their destinations due to the low frequencies employed. The level of atmospheric noise called static was high, and to reasonably develop the traffic capacity capability of a circuit it became necessary to locate the receiving terminal in as northerly a latitude as possible where such noise was less. High speed and multichannel signal methods were out of the question because they could not be adapted to the low frequencies used. Also, the number of circuits which could be operated simultaneously was very limited because of the great range of the transmissions and the rather small width of the total useful frequency band which was of the order of 85 kc or about only 20 kc for the really long distance requirements.

Utilization of the newly discovered properties of high-frequency wave propagation was accompanied by high-radiation efficiencies with resulting low powers and smaller antennas, atmospheric noise was much less, radiations could be sharply beamed, the geographic location of the receiver was relatively unimportant, high signalling speeds and multichannel operation were feasible, and there was some 15,000 to 20,000 kc of frequency space available. Nevertheless the new method was not without drawbacks. It became almost too easy to establish circuits with the result that use of the common transmission medium became a temptation leading to congestion and the creation of unjustified interferences. From two to five different frequencies were needed to maintain continuous communication over any single route to care for diurnal and 11-year solar cycle variations. Signals were adversely affected by magnetic storm and other sporadic solar disturbances causing service interruptions. Multipath effects introduced distortion which is particularly troublesome on radio telephone circuits. However, the advantages far outweighed the disadvantages, and since about 1927 the previously used methods and facilities have vanished except in a few exceptional instances where an alternate facility is maintained for use when high-frequency circuits are incapacitated by solar disturbances.

The early methods of operating international radiotelegraph circuits corresponded to those then prevalent with the older systems. Morse signals were sent by hand or from punched tape and received by sound by an operator who wrote the message out while listening. After the first World War it gradually became the practice to record the Morse characters as an undulating line on paper tape as had been the custom with submarine cable telegraphy, the operator writing out the message by visual reading of the tape. When high frequencies came into use recording on tape became a necessity in order to cope with high transmission speeds. Enough operators would simultaneously prepare transmitting tape and transcribe receiving tape to maintain the required circuit speed. By 1940 this had become a standard method although telegraph printers had been successfully adapted to a number of important circuits. Also a few long-distance radio circuits had adopted multichannel operation. Following World War II, largely as the result of the accelerated technical development of the art during the preceeding years, the major international systems adopted the use of printers and multiple-channel operation generally thereby reducing the need for skilled operators, fixing channel speeds to be within the capability of one person per channel, and enabling the automatic relaying or forwarding of messages by either regenerating the signal at the relay point directly or regeneration by punching tape and retransmitting when convenient. At the present time these methods have been refined to the point where an error caused by transmission difficulties can be detected automatically and is not recorded at the receiving terminal, which action causes special signals to be transmitted back to the sending terminal initiating the retransmission of the correct character which is then printed at the receiving end in a normal manner, all without human intervention and without disconfiguration of the appearance of the typed message.

The past 25 years are replete with many technical developments. Intensive studies of the propagation of these high-frequency waves and the mechanisms which influence the ionosphere without which these waves would not be bent back to the earth's surface have led to the ability to predict propagation conditions with considerable success. Specialized antenna systems have been made available permitting the selection from among several of the strongest arriving rays of radio waves which assist to reduce distortion and fading of the received signals. The single side-band system was produced which reduced signal distortion and improved the efficiency of use of the radio spectrum by transmitting only the frequency components actually necessary to adequately convey the intelligence to be communicated. This method also conserved power through this suppression of nonessential radiations. The development of multichannel single side-band radiotelephone equipment made it possible to greatly conserve frequency space by providing two circuits in the space previously required for one double side-band circuit and three or four circuits in only slightly more space. The use of quartz crystals to control the frequencies of the transmitting stations and tuning of receiving stations also contributed to improved operation and more efficient utilization of the frequency spectrum. Methods were developed for the transmission of pictures and a widespread radiophoto service has become available. Many other technical advances have gradually come into use.

During these years of attaining greater perfection through the evolutionary processes of invention, experimentation, and experiences gained from operating radio communication circuits, world-wide systems were being developed slowly and several distinct eras can be recognized.

The 1901 discovery that the Atlantic could be spanned by spark-produced radio waves led to the first international radiotelegraph service in 1907. The Marconi interests eventually built spark-type systems for spanning both the Atlantic and Pacific Oceans during the period 1912 to 1917, and plans were offered for establishing a British Empire system, which, however, did not materialize. They opened a commercial circuit across the Pacific in 1915, but the World War interfered with the establishment of a transatlantic service. A German company established a circuit between Germany and the United States in 1911 using a quenched-spark system, but the signals were not sufficiently reliable to maintain an adequate service. Thus the spark method lasted only some 10 years in this field.

In 1912 the Federal Telegraph Company opened a service from San Francisco to Honolulu using the arc generator method. The United States Navy adopted this system and established a chain reaching from Washington to Panama and Manila using powers as high as 500 kw. Stations were constructed in Italy, France, Ireland, and Java and during World War I several transoceanic circuits were in service using this system. The largest arc-type station was completed in France in 1920 having a power of 1,000 kw with an antenna supported on eight towers 820 feet high. This method also had a practical life of about 10 years.

The first transatlantic commercial system using high-frequency generators was constructed by a German contracting firm for a French company and employed the Goldschmidt machine. Service started in 1914 between the United States and Germany, the contractor seizing the plant at the United States terminal since France was at war with Germany. A second such circuit established by another German company commenced operations in 1915 using the German Arco-type machine, supplanting the spark-type system established in 1911. An alternator system developed by Alexanderson of the General Electric Company was put into regular use in the United States in 1917. Other types of alternators were developed in France and Japan. Following World War I services were established across the Atlantic, the Pacific, with Buenos Aires, Indo-China and other places using alternators. This system can be said also to have had a life of 10 years as no more significant additions were made after 1924.

The first high-frequency international services to attract compelling attention were those created by the Marconi short-wave beam system. Having been left behind when arcs and alternators displaced the spark generators, the Marconi Company was quick to set the pace in in this new technique and services were opened between England, Canada, the United States, India, South Africa, and Australia in 1926 and 1927. The capital investments for these were about one half of those needed for the previously used systems, the performance was superb in comparison, and high transmission speeds were available. The competition with the submarine cable companies thus created rapidly led to a merger bringing about the British international telegraph monopoly (except for services with European continental countries, which were operated by the Post Office), of Cable and Wireless, Ltd., which recently became nationalized. Simultaneously the Radio Corporation of America and the Transradio Company of Germany installed high-frequency facilities as auxiliaries to support their low-frequency circuits.

Very shortly after the establishment of low-frequency radiotelephone service across the Atlantic in 1927 supplemental high-frequency channels were added. The first high-frequency circuit between the United States and England, opened in June 1928 by the American Telephone and Telegraph Company and the British Post Office, was followed later that year by a circuit from Germany to Buenos Aires. In 1929 the Dutch established one to Bandoeng, another was added between New York and London, and one was established from Madrid to Buenos Aires. Other circuits followed rapidly between North and South America, between American and European points and elsewhere, so that by the outbreak of World War II the world was covered by a multitude of radiotelephone circuits. Subsequent growth makes it now possible for almost every country to have international telephone service available to it.

The seeming extensiveness of the frequency spectrum space available combined with the low cost and ease with which high-frequency stations could be constructed brought about a tremendous expansion of radiotelegraph and later radiotelephone stations. Great world-wide radio communication centers became established in all important countries. Internal systems for large countries like the USSR and Brazil sprang into being. Systems for military purposes became very extensive. By the time of World War II space in the high-frequency part of the radio spectrum became more difficult to find, newer services such as for broadcasting and aviation had to be accommodated, and frequent international conferences to deal with improved technical standards, operating practices and the allocation of frequencies took place. Since World War II and following the protracted International Telecommunication Conference at Atlantic City in 1947 the matter of radio-frequency allocations and radio-station assignments has become an involved international activity engaging the continuous attention of many persons including those comprising the permanent staffs of the International Telecommunications Union located at Geneva.

The existing international services

cover every facet of telecommunication need. In addition to normal message and telephone service other requirements are accommodated such as for aviation safety and flight control purposes, collection of weather data, picture and facsimile transmission, broadcast program transmissions, and leased channel services. Radio has provided the means for many of these services. Prior to the entrance of radio methods in the international telegraph field about 4 decades ago, the total mileage of submarine cables in service increased approximately 4 per cent annually over a period of many decades. This rate dropped to about 2 per cent during the time international radio communication was developing, and since 1930 the mileage of abandoned cables has exceeded that of new cables. Since the advent of the high-frequency method about that time, radio circuits have provided for practically all of the growth in the international telegraph field.

The high-frequency method has now held the field for some 25 years and there seems little likelihood of its soon being displaced. Superhigh frequency radio relay chains have been highly developed in recent years and now provide many very high quality communication routes over land. In fact, such a system provides a virtual highway having a tremendous traffic capacity for all conceivable forms of telecommunication service including television program transmission. Until recently there has appeared to be no practical way to apply such chains for intercontinental purposes since ultrahigh frequency waves have been believed to have only short range reliable propagation ability and locations for the numerous relay stations needed on this basis do not exist. Recent discoveries, however, give some basis for hope that with further development the means may be found to use these waves reliably over considerably greater distances. A worldwide system could be visualized connecting the hemispheres via the Behring Strait as once was planned for a wire line between New York and Moscow and perhaps some day African and European as well as North and South American areas will be so interconnected. However, economic and political factors constitute formidable barriers. Relay or reflection methods using the moon or an artificial satellite to provide routes for ultrahigh frequency transmissions are interesting possibilities but, unfortunately, the lunar cycle interferes, and how to establish and control platforms in space is yet to be learned. The ease with which the high-frequency method enables any country to communicate with almost any other virtually without regard for distance has made radio communication a national asset and in this time of growing nationalism it is unlikely that any new method that takes this advantage away could be established successfully. Future developments, therefore, are more likely to be in the direction of organizing the flow of communications over routes where the troublesome effects of natural disturbances to wave propagation will be at a minimum and improving automatic operational methods including the application of modern switching methods as have been developed in the wire transmission field.

International radio communication has progressed from the detection of a rudimentary signal sent across an ocean to the present extensive and ramified services it provides in the brief period of 50 years. A basic stimulus has been the innate desire of mankind to converse. With that stimulus it can be predicted that new methods and systems will continue to evolve but our present powers of visualization are inadequate to foretell accurately the forms that they will take.

It is to be hoped that improved communication facilities in the future will not be artificially restricted or controlled and that world-wide freedom to converse will provide a further incentive to develop better technical instrumentalities.

No Discussion

The Impress of Enterprise upon Telegraphy's First Century

I. S. COGGESHALL
FELLOW AIEE

Synopsis: The relative merits of private versus government ownership and operation are examined in the field of electrical communication, advantage being taken of evidence covering approximately a century, distinguishing American experience in private corporations from that of Europe in state administrations. Principal attention is focussed on telegraphy, as oldest of the electrical industries, and on those features which are of most interest to engineers and management. It is concluded that while particular forms of organization do exert an appreciable effect upon discovery and invention, their most obvious impress has been upon the respective degrees to which they have exploited invention to maximum public benefit. The disparity in effect of American and European casts of enterprise upon world progress in communication is examined pragmatically, and also analytically as to cause. Social trends leading to changes in conformations as they now exist are assessed from the viewpoint of their possible bearing upon American telecommunication.

MUCH of America's industrial preeminence is founded upon technical discoveries, inventions, engineering design, and experiment into all of which the labors of trained engineers are absorbed. Added to the large number of graduates entering those specialties each year in the field of electrical engineering, it has been estimated[1] that there is a loading of 50 per cent representing other engineers employed in administration and management.

Those of the latter who are in the private communication industries are in the tradition not alone of Morse, Bell, and Marconi, but of Taylor, the Gilbreths, and Gantt. Perforce, they ride the swinging pendulum in the traverse of its historic arc, from sweated labor at one extreme to featherbedding at the other, valiantly trying to dampen its oscillation in the interest of industrial amity between the thousands of people who own the properties and the thousands who operate them for mutual profit. Through mechanization, and especially through organization of the work, they aim to achieve a maximum of output for a minimum of

---

Paper 52-257, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Program Committee for presentation at the Centennial of Engineering—1952, Chicago, Ill., September 10–12, 1952. Manuscript submitted July 1, 1952; made available for printing July 24, 1952.

I. S. Coggeshall is with the Western Union Telegraph Company, New York, N. Y.

time, toil, and nervous energy expended by the workers; and with what wisdom is vouchsafed to them, they help bring about a fair division of the resulting wealth among users, employees, government, and owners of the means of production.

Such engineers-in-management are attentive to the movements[2] which today are resulting in the shifting of laboratories, offices, and factories to outright government ownership, various degrees of control, or financial support by the state. Other groupings of members of our profession are equally attentive: engineers in consulting firms retained to construct and operate public works; others,[3] motivated by a desire for the public weal in perhaps an even more direct way, who enter the government service. Considering these varying shades of self-interest and the interplay of philosophies and economics over a long period of time which has brought them about, one who bears in mind the good of the profession as a whole is unlikely[4] to assume a didactic, unequivocal position in favor of private, and against public, operation (or vice-versa) of all electrical utilities having a wide base of public service.

Yet there would appear to be merit in examining the record of a century of telegraphy under the different forms of enterprise to determine whether it throws light on the question: Due to the unique requirements of organizing telecommunication for speed,[5] are there inherent advantages in the private form of operation which, on the record of telegraphy, would dictate their exceptional continuance in this country, even in face of a gradual drift toward government ownership of utilities?

Private Operation of Telegraphs in the United States

The United States alone of the principal nations of the world operates its internal and external communication under private auspices. Morse built his first proprietary line in 1844 under a Congressional grant, operating it for nearly a year from quarters first in the Capitol then in the Post Office, free of charge to the public, with government aid. It then went under Post Office operation with a public tariff, was next leased to private operation, and finally sold to private owners—all[6] in the space of 3 years. Congress had decided that private ownership was more in keeping with the pioneer tradition of our country.

Western Union was incorporated in 1851, assuming its present name in 1856. Within 20 years it had become the largest corporation in the United States,[7] having absorbed a dozen families of some 500 competing companies[8] of limited geographical spans. Until co-ordinated, they had been incapable of rendering acceptable origin-to-destination service. Postal Telegraph was formed in 1881 with name and policies slanted toward sale to the government. The plan failed but Postal became a formidable competitor of Western Union until the latter purchased Postal in 1943. Meanwhile, Western Union temporarily (1909–1914) had been a subsidiary of the American Telephone and Telegraph Company, and Postal (1928–1943) a subsidiary of the International Telephone and Telegraph Company.

Domestically, the telegraph company today is in what economists[9,10] had long held to be its proper role to render service—a "natural monopoly of organization," with unit costs decreasing with increasing loads. However, it competes for custom with two like economic entities, the telephone company and the Post Office, and is regulated by the Federal Communications Commission. The company's size, scope of operations, equipment and operating methods, and business trends on this 100th anniversary of engineering have been described officially.[11,12] AIEE, during the centennial of telegraphy in 1944, published monumental papers, which are still pertinent, on telegraph,[13] telephone,[14] and cable[15] companies' telegraph operations under private management.

Threats of Government Ownership

With practically all the world except the United States and Canada combining telecommunication with postal facilities, there have been recurrent demands for government ownership of telegraphs here. A flare-up in the Lincoln administration gave the government an option, which it never exercised, to purchase the lines. In 1873 a new corporation was to acquire all lines, terminate them in post offices, and they were to be operated by the government.[16] This also fell through. John Wanamaker, Postmaster General under President Harrison, also advocated government ownership but obtained no action.

In the middle of World War I, a double crisis arose as the result of labor difficulties in the telegraph system and a shortage of telephone instruments. The government took over and the Postmaster General operated both telegraphs and telephones for a little less than a year, with considerable financial loss and an unwelcome diversion of government attention during a major war. The properties were returned to their owners and have remained in their hands.

Upon the basis of powers conferred upon the President in the Communications Act of 1934,[17] a different course could be followed when World War II broke out. Instead of taking over telecommunication properties, the government obtained the support it desired by co-operation, turning to the companies constantly for aid[18] and never having to exert powers conferred by the Act. At the termination of the war there was no commandeered property to turn back, no claims to be settled through litigation, and no inheritance by the government of nonproductive problems such as it had had to meet in the aftermath of the previous war. It had been demonstrated that government ownership of the facilities of electrical communication had nothing to offer which private enterprise[19] could not produce better in the prosecution of a major war.

Present Government Policies

Subsequently to the 1918 instance, there has been no occasion for resort, in communication, to the now frequently used device of temporarily donning government ownership like a greatcoat to weather a season of labor strife. The President's Communications Policy Board in 1951 declared for the Executive branch of the government the policy:[20] "The United States almost alone among the nations of the world, relies on privately owned telecommunications companies to play the principal part in the country's telecommunications system. It should continue to be the policy of the United States Government to encourage and promote the health of the privately owned companies as a vital national asset."

Senator McFarland has summarized,[21] unchallenged, what may be taken as the point of view of the Legislative branch: "While few realize it, the future of our kind of life is involved in the kind of communications policies we have. All over the world, the avenues of communications are in the hands of the state. Only in the United States do we have a free, privately owned system. We want to keep it that way."

Government Ownership of Telegraphs Abroad

In France, the government has always operated the internal electric telegraph system, as it did the predecessor semaphores.[22] Both were looked upon as prerogatives of the Crown and both were arrogated to it for military intelligence and news for some years before they were thrown open for private correspondence.[23] The French liasion of telegraphs with the mails was then copied by all European countries, the United Kingdom being last to fall in line. France also assumed leadership in effecting treaties and telegraph conventions among the European states, fundamental to compatibility of communication across national borders. Nevertheless, the French Administration of Posts, Telegraphs, and Telephones has, since about 1910, lagged far behind developments in the United States. The telegraphs consistently operate at a deficit which is supported by earnings in the posts and telephone branches.

In England, telegraphy began as a private enterprise in parallel with development of telegraphs in the United States. For a period of about 22 years, until after the first transatlantic cables were in use, the pattern of private ownership prevailed. Two strong companies and two smaller ones evolved; there were consolidations and rate wars, as in this country.

Commencing in 1854, pressures in favor of copying the French Government system of country-wide flat rates and of obtaining relief from the overcompetitive excesses of the private companies and the railways' telegraphs led to demands for government ownership.[24] In 1868–1869, all the inland telegraph companies and the commercial telegraph business of the railways were made an operational monopoly of the Postmaster General. The system has been so operated to the present time.

Earnings and Deficits

One is obliged to examine Parlimentary Debates[25] in addition to earnings and deficit figures of government reports to discover the situation underlying the accounts of the inland telegraphs. There is lack of uniformity of presenting the figures from year to year; expenditures of other departments and the railways for telegraph service are befogged with value of services rendered other departments, and, as is characteristic of government accounts the world over, net revenues may be increased as required or a deficit changed to a surplus by ignoring expenditures for buildings, sites, plant extensions, and pensions. Despite these maskings and favored circumstances, British telegraph deficits have been the rule,[26] with only scattered years of net revenues. At the present time, operations are running at a deficit.

By way of contrast with Britain and France, Western Union in the 96-year interval 1857–1952 paid dividends in every year but 11, and from 1873 to 1951 failed in only 5 years to show net income from current operations. While doing so, it carried every year what to Post Office accounting[27] would appear as the added burdens of paying taxes in every state and county, paying office rents, depreciation and insurance on property, and pensions, and paying interest on its loans and bonds without resort to transfer of items to other than its own accounts; and absorbed all losses without recourse to taxation.

Europe's Overseas Telegraphs

The government telegraph and telephone administrations of Europe, including England and France, maintain a government monopoly of wires and cables crossing state borders as part of the intra-European system. Many of the countries extended this control to all international circuits, but England, France, Germany, and Italy have permitted private and quasi-government companies to operate cables and radio circuits abroad, and other European countries have countenanced private cable and radio ownership.

Two American[28] transatlantic cable companies maintain competing offices in Paris, Le Havre, Belgium, the Netherlands, and in London and many other cities of the United Kingdom. Quasi-governmental Radio-France is the Radio Corporation of America (RCA) Communications' French correspondent; the French Administration is Mackay Radio's, both American radio companies correspond in England with the British merger company, Cable and Wireless, which was formerly privately owned, but is now nationalized.

The British Merger

The British merger of cables and radiotelegraphs took place in 1928–1929 under governmental control by Britain and the Dominions. The managerial function continued to reside in a general manager but under a handicap somewhat caustically described by Newcomb Carlton, an engineer and president of Western Union, before a Senate committee[29] in 1930: "I know those people intimately. We have been doing business with them for 20 years.... There is a representative from each of the colonies and one from the British Treasury.... They have no responsibilities and no executive authority. If the operating company of the Merger wishes to lay a cable... they have to go to this national commission for a decision... In respect of any important executive act... they must first go through this ponderous national machine. All very nice fellows. But they have the authority without the responsibility. And that is always a deadly thing in business.... A more creaking, ponderous, awkward business setup has never come under my observation. And as for fearing it! I wish my competitor no greater luck than to be hooked up on lines similar to the British Merger. The British Merger... will work satisfactorily only when it is changed."

The quotation is of significance in showing the impatience of an engineer in management with the circumlocutions of government in industry. One does not criticize one's friends unless the issues lie deep—as they do in the matter of the superiority of free enterprise in getting work done.

Carlton's criticism, if heard in London, went unheeded. In 1936, the Post Office's overseas beam radio property was brought into Cable and Wireless, and along with it more home government control. Finally, in the interval 1946–1950, plans were made and consummated for the company's nationalization, with the British Government the sole shareholder. In Parliament, 20-year-old echoes are heard today[30] of Carlton's diagnosis: "Has not worked out as hoped... Users have grounds for dissatisfaction... Alarming tendency for customers to shift to the American companies... Eight autonomous parts... Nothing that can be changed... Now no counterpart to the Chairman of Cable and Wireless from 1926 to 1946... No central direction of the Commonwealth system... Lack of flexibility and power to take hour-by-hour decisions vital to a competitive international system."

There are usually two sides in legislative debate. Were there no amelioration of the symptoms described, traffic movement under government had reached a state of distress.

Progress as a Function of Technology

In seeking the causes of governmental deficiencies in communication, questions concerning adequate technological competency naturally arise.

Are engineers employed by foreign governments technically incompetent? All evidence points to the contrary. Their publications are on the highest plane, as are their papers presented before technical societies and especially the record of their discussion. Their visiting emissaries almost uniformly are impressive, their engineers-in-chief often profound.

Is there a paucity of inventors among them? Decidedly not. In electrical communication this has been true from Samuel Morse to Sir Watson-Watt. As ideas and inventions have been lobbed from Wheatstone to House to Kelvin to Bell to Belin to Pupin to Marconi to De Forest to Eccles to Zworykin, national boundaries have been but nets over which the shuttlecocks flew. No nation has a corner on inventiveness or ingenuity. No system of enterprise maintains its lead in communication by virtue of invention alone.

Do they lack intellectual curiosity? By no means. A steady stream of foreign government communication engineers visits our shores every year to see how we do it. They see only a little more in the way of unit apparatus than what they were prepared to see; in many cases they have like units in their laboratories or set up in their operating rooms. They are exceedingly curious (be it said to their credit) for reasons not connected with volts, cycles, or decibels, as to why it is that in America they see these units row on row, the decks having been cleared of antique clutter, working in profitable operation and giving the public a superior service.

Is American equipment unavailable to them? To the contrary, most American equipment is regularly on sale abroad, usually priced low enough so that if used under circumstances akin to those found under private enterprise it would yield an investment return in a reasonable time. The 5-unit code, the multiplex, the stop-start teleprinter, facsimile, and the reperforator were all involved in round-trip European-American odysseys.

Have they no laboratories to develop their own apparatus? Certainly they have. But in the comparable peace year 1935 when the Western Union general engineering and research staff numbered 181 qualified engineers (it is double that figure now), the Dollis Hill Research Station staff of the Inland Telegraphs comprehended 112 in the engineering classifications, indicating a 60-per-cent lead in the measure of private industry's support of telegraph engineering. As a matter of fact, the disparity was considerably greater because properly included in an American count would be many Bell System engineers working on telegraphy; also, cognizance would be taken of the scope of the Post Office's researches as inclusive of telephony as well as telegraphy. In comparing Dollis Hill with the Bell Laboratories in the United States, a Post Office author[31] concluded that, measured in relative terms of expenditures per year per telephone subscriber, American enterprise in the middle 1930's was spending a dollar against the British equivalent of 19 cents.

Is there a missing, priceless ingredient? Competent observers say yes; the something lacking is spirit. Inventiveness is regenerated by tangible returns. It has thrived best where enterprise has been ruthless in replacing still-useful apparatus; where capital has been promptly placed to effect changes dictated by economics, efficiency, and service improvements; where this has been done on a widespread scale, not in piecemeal, faltering fashion; and where no sentiment has been allowed to stand in the way of progress except concern for the people working the circuits. This is the record of American enterprise, in which the span of one man's career has witnessed the scrapping of Morse to install multiplex, the destruction of all but specialized multiplexes in the process of installing teleprinters, and the replacing of teleprinters by facsimile in large areas of customers' connections; where page-printers have been scrapped for tape-printers, open wire lines disposed of in favor of carrier facilities, and message belt-conveyor systems junked in the introduction of reperforators.

On the contrary, inventiveness has thrived least where ideas languish—not because they are not good but because they are not "bought"; where the heavy hand of government inertia has smothered all sparks of ambition under a wet blanket of politics, oblivious to gains achievable through standardized equipment and methods, the specter of past years' deficits, incompatibility of proposals with existing plant, and a straitjacket of international agreements and regulations; where the engineer-in-chief has met with repeated rebuffs as he has gone to the exchequer year by year with his capital budget, only to find his requirements in conflict with causes latest in the public need or fancy, and appropriations going to the politician with the loudest voice instead of to the one with the best case.

Is inaction inherent under public ownership? Probably. The profit motive provides a strong drive and insures a continuity of stimulus, from the top down, to create of a laboratory a continuous source of income and improvement. By aggressive methods of merchandising which governments seldom learn, new discoveries are translated without delay into the new revenue which accompanies their public acceptance. The profit system has the virtue of quickly testing the public's sense of relative good because buyers have to allocate pocket money to back their preferences. In the testing process, industrial management, streamlined for quick decisions, adapts its program and service to its customers' expressed interest and finds, in the process, that enthusiasm is as vividly communicable to the engineering staff as to accounting and sales. The spirit thus built up to produce more ideas and resulting better products and services cannot be felt in full force, or harnessed, except under capitalism, for consumer acceptance is a more democratic and sensitive touchstone of the worth of a good or service than are all the dictates of a planned economy, planned though it be by an elected administration.

Progress as a Function of Organization

Creed, a noted telegraph equipment manufacturer in England, was once pondering the frequency with which the British Post Office changed its engineers-in-chief.[32] "When you have got 'well in' with the Engineer-in-Chief, they change him," he observed. "As soon as you have put your views fully before the Secretary they change him and make him a director of the Turkish Bank or something like that." It is typical of government operation or control of utilities to promote top-ranking officers, conversant with government finance or the law, horizontally into positions where it is assumed they can learn a new business. American private enterprise typically fills its top positions by vertical promotion of men whom internal competition has demonstrated to be best fitted to accept responsibility.[33] Engineers ascending this ladder have the advantage of being grounded in the science upon which electrical communication rests, and thus appreciative of the physical possibilities and limitations of the business.

ENGINEERS IN COMMUNICATION MANAGEMENT

So far away from any tradition of engineering capability in management had the British Post Office kept, that in 1931 a Royal Commission[34] on the Civil Service "expressed surprise that no engineer had ever been promoted to an administrative post in the Post Office!"

An engineer was president of Western Union at the time; since his retirement another president has been drawn from engineering. An expected high concentration of engineers in key posts under the vice-presidents of the research and development and the plant and engineering departments has been supplemented for years throughout the organization. Engineers are represented on the board of directors; the assistant to the president is an engineer; so is the vice-president in charge of operations, and his staff and eschelons of the nation-wide line organization are weighted with graduate engineering talent. The general purchasing agent is an engineer, as is the director of the contract department. The valuation and plant accounting groups are dominated by engineers. The vice-president in charge of employee relations, himself an engineer, employs engineers on his staff and as his regional negotiators. An engineer heads the employees' benefit activity. The staff functions of planning and budgeting, both capital and operating, landlines and cable, are carried on chiefly by engineers. Several of the larger telegraph offices throughout the country are headed by engineers. Members of the profession are found throughout the field organization occupied with supervision of construction of lines, installation of offices and apparatus, maintenance, the electrical regulation of circuits, the routing of traffic, determination of operating methods and standards of performance, and supervision of the operating forces.

All this activity was well under way at the time of the Royal Commission's report. Because what is said regarding engineers in the telegraph company's management could be repeated faithfully for other communication properties under private operation, a valid conclusion might be drawn, on this centennial of engineering in America, that the practices of American private enterprise and of a typical European government communication administration, in respect of exploitation of engineering management possibilities, are totally antithetical.

A Government Appraisal of Private Communication

In 1928 the British Post Office appointed the Simon Commission to visit the United States to determine reasons for financial prosperity of American telegraphs during one of the recurrent lows in government operation. The Commission reported[35] as news, because the observations did not fit government practices that:

1. American companies were pioneers in the use of machine telegraphy.
2. Western Union had "ever been on the lookout for a good and reliable system." (This was inaccurate. It was the Post Office which had remained on the lookout. The very machine which had been standardized here in 1914 from British designs,[36] and installed throughout the United States in the intervening 14 years,[37] had been installed in 1915 between London and Manchester[38] by the Post Office and virtually shelved there, while private enterprise here made it the foundation stone of its own prosperity.)
3. Generous use was made of labor-saving devices. (One result of the Simon Commission investigation was the desultory copying of American devices on a limited scale.[39,40])
4. "Americans build on Monday, install on Tuesday, and scrap on Wednesday," but British ideas of poor workmanship in American machines "must be taken with a large pinch of salt."
5. United States circuits had been engineered for "cast-iron stability."
6. High circuit outputs depend upon circuit stability and high standards of maintenance.
7. The Post Office had been hampered by want of spare (reserve) apparatus in the quantities furnished in American practice.
8. European practice suffered from too many different types of apparatus. (This observation accords with the museum-like nature of the Central Telegraph Office in London before its destruction in an incendiary raid in World War II. It had the "very latest and one of everything else." The French Administration office in Paris is similar.)
9. Frequent changes in British overhead personnel lower the standard of general efficiency.
10. American practice benefits by specialization.
11. Great Britain was tied to an "average" system of low operator output with rotation around the office, while Americans placed the best operators on the most heavily loaded circuits and kept them there.
12. Pay should be proportional to circuit speeds.
13. United States operators were carefully trained to observe operating routines.
14. American officials do not permit traffic to "fall away without any attempt to stop the rot." (Even today, the Post Office does little canvassing of patrons, in contrast to the practice of telephone and telegraph companies in the United States.)
15. Private enterprise "doesn't wait for the traffic to come to them, they go out and seek it."
16. Americans keep the business in the public eye.
17. They constantly seek new uses to which the system can be put.
18. A government system wastes effort[41,42] instead of "encouraging business people to make more and more use of the service."
19. In America, the telegraph system is free from political influence.
20. The telegraph system in the United Kingdom "has been a political shuttlecock ever since 1868" (the year it changed hands from private to government ownership.)
21. Private companies are permitted to levy charges sufficient to ensure a reasonable profit.
22. "There are no pessimists in the telegraph business in America."

The Tempo and Perquisites of Government Employment

The impeccable nature of the Simon Committee assures the fairness of the comparisons drawn. The findings give support to general knowledge that government jobs the world over are not particularly demanding, that force assignments are not tight, supervision is almost nonexistent by standards of private industry, time clocks are rare, the gait leisurely. The rolling-in of the tea lorries for afternoon tea in the Central Telegraph Office in London is not anachronistic in a government office setting. Low individual output,[43] time paid for but not worked in the form of cumulative sick-leave grants not contingent upon actual sickness, as well as liberal vacations and early retirements from service, make government communication labor under Civil Service expensive even at lower hourly wages.

Because government promotions and slotted pay schedules rest so solidly on seniority, government service in the United States and elsewhere fails to provide measuring sticks of efficiency upon which progressive private management relies to make its promotions. While a weeding out and seasoning process[45] goes on continually in industry, the motivation of government service is tenure, representing at best a dilution of enthusiasm by concern with security and at worst a consuming of the lower eschelons by sloth, the ranks being occupied by "untouchables" protected by Civil Service legal red tape and appeal machinery from being fired even for cause[46,47] and coasting towards an early pension.

It would be a great pity if the private communication industries were to bargain themselves into the same box, for preoccupation with security is an imported European tenet indigenous to government service there and until recent years given short shrift in the United States. Ideas which sprang from ancient government doles, and pittance pensions for government telegraphists and telephonists, among others, have come overseas and are finding their way into the spirit of some of our more modern labor contracts.

THE CHANGING INCENTIVES OF PRIVATE EMPLOYMENT

By and large the American telegrapher is happiest at work when he is kept idle least. If he excels in his skills he would prefer to have his working contract so drafted that his co-workers would have to acknowledge them; and if his superiority could be turned to dollar account, so much the better.

Any tendency within private industry to adopt the more leisurely tempo and norms of the government services helps destroy the lead of American over European productivity; and, by making more natural and likely a transition to government operation in the United States, jeopardizes maximal industrial support of armed might to resist aggression. In some respects bureaucracy and labor leadership have dallied with this retrograde movement in ways transparent to most engineers-in-management.

Government officers and employees in Washington regulate increasingly large sectors of industry, frequently having to weigh conflicting testimony on working conditions submitted by companies and unions. The criteria of optimum hours, conditions of work, time off with pay, and pensions, which persons in the government apply, must subconsciously be those of government offices to whose ways they are accustomed. The same is true of the "public" members of arbitration panels, to whom the cloistered environment of institutions of learning and the halls of government may furnish a poor standard for gauging life in a communication operating room or factory.

Labor leaders who have come up through working ranks should know better, but their necessarily continuous bid for power involves widening, between men and management, a psychical breach which only they must span. There are no better ways to do this than to restrain the flow of ideas, up and down, between managers and employees and to activate an organization of shop stewards to confront the foremen at every turn, in the guise of advocates of workers, with captious grievances against the shop rules by which the boss gets the establishment's work done. By thus dividing allegiance and authority, and by emasculating rules promoting effective operation, production can be brought down to levels no longer characteristic of America's enterprise nor adequate for exercise of her might. In the process, employees' attitudes are conditioned to embrace government operation of industry. Both landline and cable unions of the telegraph company have declared themselves, at one time or another, in favor of government ownership of telegraphs.

Conclusion

To a peculiar degree, electrical communication requires co-ordination of both ends of its circuits, preferably by organic control. In other industries, when co-ordination drags its feet it is merely exasperating; in communication it is fatal. The type of private organization adopted by the telephone and telegraph companies in the United States gives the service product one of its essential qualities—speed measured in seconds and minutes.

The history of telegraph enterprise in this country in the period 1844–1890 was a history not alone of financial consolidations of smaller companies and extensions of physical coverage but a history of what would now be called industrial engineering, with problems of standardization, layout, and traffic routing dictated by the nature of the machine spread out, as it was, all over the map. Later the telephone system grew up here in a similar manner, with its circuits responsive to unity of direction.

In Europe the problem of collaboration was complicated by the jigsaw puzzle of separate sovereignties, by language differences, and by the fact of government ownership. A series of conventions was entered into, having the force of treaties, and requiring the indulgence of the Foreign Offices to change. Theoretically the telecommunication administrations are free to act within their own borders; actually they are restrained in international shackles. Characteristically, political bureaucracy has evolved an organization whose diagram of committees blossoms downward in three colors like the burst of a rocket. Subcommittees of these Comités Consultatifs Internationals (CCI) meet and assimilate engineering, traffic, and tariff data in order to make recommendations for adoption, first by the CCI committees on telegraphs (CCIT), telephones (CCIF), and radio (CCIR); then by formal conferences at which envoys affix their signatures to documents, contingent upon later legislative ratification or promulgation by decree. This process of almost endless discussion may hold up for years needed changes in equipment or operating procedures which under private management in the United States are effected by turning the handle of a mimeograph or by an exchange of letters with a correspondent carrier.

Thus, somewhat perforce, American communication has grown up the agile way and European the cumbersome. The telephone system in the United States is copied throughout the world but is unapproached anywhere save where the systems have been provided by American engineering and suppliers. The American telegraph system, with its 5-unit printers, has spread throughout the world over American-operated cable and radio-telegraph circuits. The cable companies—Commercial, Western Union, and All America—have accomplished their successes by owning and operating their offices in foreign countries, with standardized equipment at both ends. Tropical Radio has followed a like course of ownership in Central America. RCA Communications and Mackay Radio in radiotelegraphy, the American Telephone and Telegraph Company in radiotelephony, and Globe Wireless and other American carriers in both fields of radio, have encountered Herculean tasks in securing the measure of co-ordination with foreign government administrations necessary to equip their joint circuits with apparatus meeting American standards of quality and essential to operating them under routines meeting acceptable criteria of accuracy, speed, and low unit cost.

The problems of compatibility between these systems, which have grown up here under private enterprise and abroad under aegis of the governments, are currently of moment as the result of the confluence of two developments. Fortuitously, engineering advances in cable and radio, which make it possible to interconnect telegraph users across the oceans, are occurring at a moment in history when America is entering into the details of world affairs by eleventh-hour participation in all sorts of expanded European bureaucratic activities. The United States is now signatory, for the first time, to the international telegraph regulations in consequence of which the Department of State and the Federal Communications Commission now represent in world councils the general telecommunication interests of America's private enterprise.

Our representatives are carrying a tremendous responsibility as they enter world councils with a pitifully small number of assured votes and with little but this nation's powers of leadership (which are not inconsiderable, but precarious) to gain more for our policies. They can hardly avoid foreign pressure through the CCIR, effective or not though it may prove to be, to compromise away the acknowledged supremacy of our independently developed radio systems in order to achieve compatibility with those of Europe; nor can the specious argument by analogy with CCIR be avoided that similarly, through CCIT and CCIF, we should clear the way, "in order to promote international good will," for substitution of European telegraph and telephone practices for United States practices in the United States.

However, American technology—American achievement in all fields of communication from telegraph to television—is held in high regard abroad, as one consequence of which this disquisition can be ended on a buoyant note of pride and hope. Nothing succeeds like success. As a tribute to our success the questions concerned with communication engineering which heretofore have been settled forthrightly and unilaterally by American privately owned carriers in relation to their own circuits will commend themselves to our friends across the seas as questions better not made subject, for the world's good, to the time-consuming compromises of foreign diplomacy. And as for jointly operated transoceanic circuits, we may expect, as another outcome of our corporate successes, and with our Government now contributing its support, that the flow of American influence, like the benign Gulf Stream, will be channeled strongly eastward, to the end that Europe, for all its handicaps of administration, may sometime enjoy as well as may be the fruits of private telecommunication enterprise.

References

1. What Kind of Engineers? **J. D. Ryder.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 37, 1949, page 1170.

2. Ownership in Common, **Dean Russell.** *In Brief*, The Foundation for Economic Education (Irvington-on-Hudson, N. Y.), volume 6, 1951, pages 1–20.

3. Are You Satisfied? **C. W. Carnahan.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 36, 1948, page 1451.

4. The Engineer in Private Enterprise, **E. W. Butler, B. B. Bauer.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 38, 1950, page 3.

5. Telecommunications: A Program for Progress (book), **President's Communications Policy Board.** Superintendent of Documents, Washington, D. C., 1951, pages 58, 78–79, 120.

6. Old Wires and New Waves (book), **A. F. Harlow.** Appleton-Century-Crofts, Inc., New York, N. Y., 1936, pages 84–87, 92–93, 104–07, 117–18.

7. Wiring a Continent (book), **R. L. Thompson.** Princeton University Press, Princeton, N. J., 1947, page 289.

8. Corporate History of the Western Union Telegraph Company, **J. C. Marchand.** Pages 4–19. Available at the Western Union Telegraph Company Library, New York, N. Y.

9. Principles of Economics (book), **H. R. Seager.** Henry Holt and Company, New York, N. Y., 1913, page 214.

10. Telecommunication (book), **J. M. Herring, G. C. Gross.** McGraw-Hill Book Company, Inc., New York, N. Y., 1936, pages 188–209.

11. Building America by Telegraph, **H. P. Corwith.** Convocation, Centennial of Engineering, Chicago, September 11, 1952.

12. **President's Communications Policy Board,** *op. cit.*, pages 63–99, 115–26, 129–30, 151–82.

13. American Telegraphy After 100 Years, **F. E. D'Humy, P. J. Howe.** *AIEE Transactions*, volume 63, 1944, pages 1014–31.

14. Telegraphy in the Bell System, **J. A. Duncan, R. D. Parker, R. E. Pierce.** *AIEE Transactions*, volume 63, 1944, pages 1032–43.

15. Developments in the Field of Cable and Radiotelegraph Communications, **Haraden Pratt, John K. Roosevelt.** *AIEE Transactions*, volume 63, 1944, pages 1044–47.

16. **A. F. Harlow,** *op. cit.*, pages 331–35, 338.

17. **J. M. Herring, G. C. Gross,** *op. cit.*, pages 409–14, 489–90.

18. Reins of War, **G. P. Oslin.** *Telegraph and Telephone Age* (New York, N. Y.), February 1946, page 10.

19. **President's Communications Policy Board,** *op. cit.*, pages 66–67, 229.

20. *Ibid.*, page 221.

21. Report of Annual Meeting. Western Union Telegraph Company, New York, N. Y., April 9, 1952, page 15.

22. **A. F. Harlow,** *op. cit.*, pages 20–24.

23. The Telegraph in Europe (book), **G. Sauer,** Paris, France, 1869, pages 10–16. Available at the Western Union Telegraph Company Library, New York, N. Y.

24. History of the British Post Office (book) **J. C. Hemmeon.** Harvard Economic Studies vii, Harvard University Press, Cambridge, Mass., 1912, pages 202 *ff*.

25. Commonwealth Telecommunications, **W. F. Deedes et al.** Parliamentary Debates, March 12, 1952 (London, England), pages 1530–37.

26. **J. C. Hemmeon,** *op cit.*, pages 215–18.

27. Government and the Moral Law, **Ben Moreell.** Jones and Laughlin Steel Corporation (Pittsburgh, Pa.), 1952, page 21.

28. **President's Communication Policy Board,** *op. cit.*, pages 151–82.

29. Hearings, **N. Carlton.** *Bill S.6*, part 11, Senate Committee on Interstate Commerce, 71st Congress, 2nd Session (Washington, D. C.), pages 1464–65.

30. **W. F. Deedes et al.,** *op. cit.*, pages 1530–33.

31. Research in the British Post Office, **B. S. Cohen.** *Journal*, Institution of Electrical Engineers (London, England), volume 75, 1934, pages 133–60.

32. The Historical Basis of Modern Printing Telegraphy, **H. H. Harrison.** *Paper Number 59*, Institution of Post Office Electrical Engineers (London, England), 1915, page 59.

33. Long-Range Business Policies: A Case Study, **K. S. McHugh.** *Bell Telephone Magazine* (New York. N. Y.), volume 31, Spring 1952, pages 28–37.

34. History of the Engineering Department, **E. C. Baker.** *Green Papers*, Post Office (London, England), Number 46, 1939, pages 42–43.

35. The American Telegraph Systems, **G. T. Archibald.** *Telegraph and Telephone Journal* (London, England), volume 15, 1929, pages 135, 158, 176.

36. Practical Aspects of Printing Telegraphy, **Donald Murray.** *Journal*, Institution of Electrical Engineers (London, England), volume 47, 1911, pages 450–529.

37. Printing Telegraph Systems, **J. H. Bell.** *AIEE Transactions*, volume 39, 1920, pages 167–230.

38. A New Type-Printing Multiplex Telegraph, **A. H. Roberts.** *Journal*, Post Office Electrical Engineers (London, England), volume 8, October 1915, pages 193–205.

39. American Influence on British Telegraph Practice, **G. T. Archibald.** *Telegraph and Telephone Journal* (London, England), volume 18, 1932, pages 75–77.

40. The Inland Telegraph Service—The Introduction of Modern Machinery and Methods, **R. P. Smith.** *Journal*, Institution of Electrical Engineers (London, England), volume 72, 1933, pages 189–228.

41. **Ben Moreell,** *op. cit.*, pages 23–25.

42. **J. M. Herring, G. C. Gross,** *op. cit.*, pages 413–14.

43. **G. T. Archibald,** *op. cit.*, page 176.

44. **W. F. Deedes et al.,** *op. cit.*, page 1532.

45. **K. S. McHugh,** *op. cit.*, page 34.

46. Why We Can't Give Taxpayers Their Money's Worth, **Charles Stevenson.** *The Readers Digest* (Pleasantville, N. Y), April 1952, pages 46–50.

47. The British State Telegraphs (book), **H. R. Meyer.** The MacMillan Company, New York, N. Y., 1907, pages 245–392.

Discussion

Jacques Treves (French Administration of Posts, Telegraphs and Telephones, Paris, France): I am not qualified to discuss the comparative merits of free enterprise and government ownership, as far as telecommunication operation is concerned, and I do not intend to do so. But, as a telephone engineer of the French Post Office, a European Government agency operating telecommunications, I felt, while listening to Mr. Coggeshall's most interesting paper, that I could not agree with some of the facts he mentioned.

If it is true that in France telegraph operates at a deficit (approximately 4 billion francs for each of the last three years), it must not be forgotten that telephone, operated by the same Government, is remunerative since the general situation in France has become more normal (15.6 billion francs in 1949, 16.4 in 1950, 20.1 in 1951).

I would not leave you under the impression that our staff works without supervision, and with the doors open. Supervision is strict on both working hours and on actual work during these hours.

Though promotions and pay depend more on seniority than they do in American private enterprise, our technicians in all ranks are well trained and efficient, and lack neither enthusiasm nor individual output. As proof of their efficiency, I would like to say they are in demand by private industry, which offers them higher pay, because private industry thinks they are good.

As another proof, I must mention the fact that, last year, I had the responsibility of installing inside telephone plants at the Supreme Headquarters Allied Powers in Europe Headquarters in Paris. It was understood that the French Post Office would be responsible for maintenance at the beginning, the maintenance being subsequently taken over entirely by United States military personnel. But, after the preliminary maintenance period was over, United States Signal Corps officers asked us to keep our switchmen working for them, saying they were so well trained and efficient that better ones would be hard to provide.

Our research stations, though numerically insufficient, have done very good work both in radio and wire transmission and in switching.

Progress is not confined to the laboratories, and the French Administration of Posts, Telegraphs and Telephones does not lag far behind in developments in all fields. I point to the fact that all subscribers in Paris dial calls directly to subscribers in such remote cities as Rouen (75 miles from Paris), Lyon (285 miles), and Toulouse (410 miles). This service operates with full reliability, has been a great success, and will be extended to other trunks.

The team of European engineers with whom I have travelled here is very much interested in seeing the American system of telecommunications, and we admire its efficiency. Thanks to American hospitality, we hope to bring home with us to France much useful information in order to improve further our telecommunication systems.

Magnetic Amplifiers of the Self-Balancing Potentiometer Type

W. A. GEYGER
MEMBER AIEE

Synopsis: This paper describes a new type of magnetic amplifier having inherent self-balancing potentiometer characteristics. By means of a special compound feedback circuit, extreme stability and the fulfillment of balance conditions in the control circuit are obtained. Positive external feedback or self-feedback produces an effectively infinite gain that is highly degenerated by negative feedback resistance-coupled into the control circuit loop. Thus, in the case of voltage, as well as current input, the balanced control circuit demands practically no energy from the input source. The effective input resistance of a self-balancing magnetic amplifier for voltage measurements is about 100 to 1,000 times higher than the actual input circuit loop resistance; the effective input resistance of a self-balancing magnetic amplifier for current measurements is about 100 to 1,000 times smaller than the actual resistance of the control windings.

The unique feature of a considerably reduced control current for the entire working range means a materially more linear relationship between the input and output quantities. Over certain ranges large variations in supply voltage, supply frequency, load resistance, and control circuit resistance have a comparatively small effect upon the "transconductance," "voltage ratio," "current ratio," or "transresistance" of self-balancing magnetic amplifiers operating under extremely stable conditions with a high over-all power gain.

For certain applications, special compensating effects can be applied by designing the degeneration network in such a way that resistance values of this network are a function of voltage, current, temperature, or time (derivative feedback effects, for example, delayed degenerative feedback).

Results of various measurements lead to the conclusion that the actual behavior of self-balancing magnetic amplifiers approximates very closely corresponding characteristics of an "ideal amplifier," just as an electronic amplifier of the compound feedback type does.

FOR many years saturable core reactors, or magnetic amplifiers,[1,2] have been applied to electrical measurements. They have long been in use for the measurement of large direct currents and high direct voltages.[3–7] Magnetic amplifiers of the balance detector type[8] based upon "external feedback" or "self-saturating" push-pull circuits have been used for the control of small reversible motors operating in connection with high-speed recording d-c potentiometers and d-c bridge networks.

In a previous paper[9] typical characteristics representing results of various measurements on magnetic amplifiers having Orthonol tape cores[10] with multilayer toroidal windings are shown, and figures for such amplifiers are quoted to indicate the performance obtained when this rectangular hysteresis-loop core material is used. These measurements were made using either indicating pointer instruments or a special d-c potentiometer device making it possible to measure the ratio between output current and input current with an absolute accuracy of ±0.02 to 0.03 per cent. The resulting empirical relationships based on experimental evidence have proved to be of considerable importance for successful design of high-performance magnetic amplifiers, especially those of the push-pull balance detector type.

Referring to the numerous kinds of magnetic-amplifier circuits utilizing feedback effects, it is important to stress the fact that these circuits generally do not have any influence upon the effective input impedance of the control circuit because the d-c feedback component flows through special feedback windings (external feedback) or in the load windings carrying unidirectional currents (internal feedback, self-feedback, or self-saturation). Therefore, the effective impedance of the control circuit is solely determined by the actual total resistance of this control circuit comprising the resistance of the control windings and a series resistor which is conventionally provided to increase the speed of response of the magnetic amplifier.

Feedback arrangements which do not influence the effective impedance of the control circuit are characterized by the fact that the winding components producing the feedback effects are inductively coupled with the control winding components. These arrangements may be termed pseudofeedback circuits, because they do not have the inherent property of influencing the actual magnitude of the input circuit impedance, as, for example, the electronic amplifier of the cathode follower type does.

Seven years ago the author showed[11] that real feedback can be applied to magnetic amplifiers, particularly for instrumentation purposes, by feeding back a voltage or current, proportional to the actual load current, resistance-coupled into the control circuit loop so that a very large amount of degeneration is obtained. In this arrangement[11] employing a magnetic amplifier of the balance detector type in combination with a phase-sensitive electronic amplifier, the actual input voltage across the control windings of the magnetic amplifier, is only about 2 per cent of the input voltage to be measured (using 98-per-cent degeneration) so that the effective input impedance of the control circuit is about 50 times higher than the actual resistance of the control windings (for example, effective control circuit resistance: 6×50=300 ohms). Similar magnetic amplifier circuits utilizing negative voltage feedback resistance-coupled into the control circuit loop, and preferably having a floating output, have been described by A. S. FitzGerald.[12–14]

This paper describes new types of magnetic amplifiers representing various advantageous modifications of the aforementioned arrangements,[11–14] employing no electronic amplifier components, and having inherent self-balancing potentiometer characteristics. By means of a special compound feedback circuit,[15] extreme stability and the fulfillment of balance conditions in the control circuit are obtained. Positive external feedback or self-feedback produces an effectively infinite gain that is highly degenerated by negative feedback resistance-coupled into the control circuit loop.

These magnetic amplifiers of the self-balancing potentiometer type have a large power gain, linear output-input characteristic, and a high speed of response. Of course, the relationship between the previously mentioned properties is such that if one property is accentuated, it is always at the expense of the other properties.

Such amplifiers are particularly suitable for extending the sensitivity of direct indicating electric instruments and ink recorders making a permanent record of small direct voltages created for instance by thermocouples, photoelectric cells, or radiation pyrometers. In the fields of telemetering and automatic control, mag-

---

Paper 52-271, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13-17, 1952. Manuscript submitted November 23, 1951; made available for printing August 8, 1952.

W. A. GEYGER is with the United States Naval Ordnance Laboratory, Silver Spring, Md.

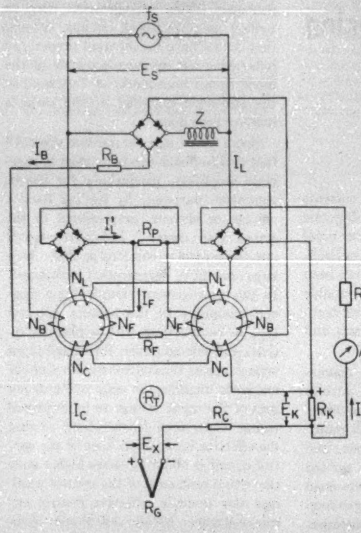


Figure 1 (left). Self-balancing magnetic amplifier of the external feedback type for voltage measurements. The transconductance of this circuit is $I_L/E_X = 1/R_K$

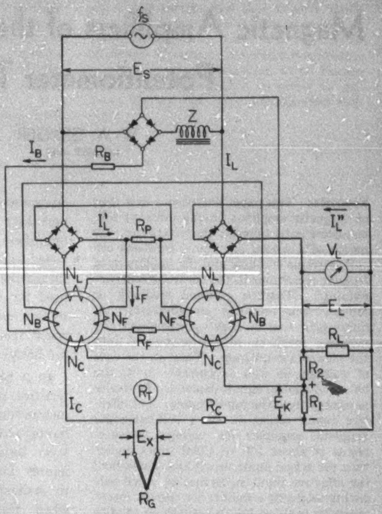


Figure 2 (right). Self-balancing magnetic amplifier of the external feedback type for voltage measurements. The voltage ratio of this circuit is $E_L/E_X = (R_1+R_2)/R_1$

netic amplifiers of the self-balancing potentiometer type undoubtedly will find numerous interesting applications because of their unique properties and distinct advantages over other arrangements based upon the use of high-sensitivity galvanometers or vacuum tubes.

Basic Magnetic Amplifier Circuits of the Self-Balancing Potentiometer Type

General

These circuits containing two separate and equally rated saturable reactor elements, preferably with nickel-iron-alloy-type cores, utilize positive external feedback or self-feedback for producing an effectively infinite gain that is highly degenerated by negative feedback resistance-coupled into the control circuit loop. Thus, in the case of voltage, as well as current input, the balanced control circuit demands practically no energy from the input source. However, it is to be noted that these simple circuits can only be applied when the voltage or current to be measured will not change polarity, because direction of the output current (polarity of the variable compensating voltage or current) is always the same.

External Feedback Circuits

Voltage-Balance Circuit. Figure 1 shows the basic circuit of a self-balancing magnetic amplifier for voltage measurements. The two saturable reactor elements have series aiding connected a-c load windings N_L, series opposing connected control windings N_C, series opposing connected feedback windings N_F, and series opposing connected bias windings N_B, so that the voltages induced in each of the d-c twin windings N_C, N_F, N_B by the fundamental wave and by the odd harmonics of the supply voltage E_S (frequency f_S) are opposed. The d-c load resistor R_L with an ammeter A_L, a series resistor R_K, and the feedback windings N_F (resistance R_F) with a regeneration control shunting resistor R_P are supplied by two separate full-wave dry-disk rectifiers with direct currents I_L' and I_L'' both representing the average value over one half-cycle of the alternating current I_L flowing through the load windings N_L. Bias windings N_B carrying the bias current I_B and reducing the "quiescent current value" of I_L are supplied by a third full-wave dry-disk rectifier with a nickel-iron-alloy-core type series impedance Z and a d-c series resistor R_B. The bias current I_B produces an additional d-c magnetization which is opposing to the d-c magnetization produced by the feedback windings N_F and independent of the control current I_C flowing through the control windings N_C.

Actual control current I_C is proportional to the difference of voltage E_X to be measured and voltage drop E_K across the resistor R_K and inversely proportional to total control circuit loop resistance $R_T = R_G+R_C+R_K$, where R_G is the internal resistance of the input source, and R_C is the total resistance of the control windings N_C. If I_C is very small in comparison with I_L, actual control current will be

$$I_C = \frac{E_X - E_K}{R_T} = \frac{E_X - I_L R_K}{R_G+R_C+R_K} \qquad (1)$$

With $E_X = 0$ the very small quiescent current value of I_L causes a voltage drop $E_K = I_L R_K$ across resistor R_K, and a very small negative control current $I_C = -I_L R_K/R_T$ is flowing through the control windings N_C. Note that this circuit assumes a conduction path through the voltage being measured so that this negative control current producing negative (degenerative) feedback can flow through the galvanically coupled control circuit loop resistance R_T.

The actual or measurable load current $I_L' = I_L'' = I_L$ is the resultant of the exciting component I_{LO}, which is practically independent of the control current I_C, and the load component I_{LC} flowing in response to changes in control current I_C. Assuming ideal rectifier performance ($I_L' = I_L'' = I_L$), the positive (regenerative) feedback ampere-turns aiding those provided by the control circuit will be $I_F N_F = I_L N_F\, R_P/(R_P+R_F)$, and the negative (degenerative) feedback ampere-turns of the control windings N_C will be $I_C N_C = I_L N_C\, R_K/R_T$.

According to the previously mentioned current-transformer character of a saturable reactor, preferably having a rectangular hysteresis-loop core material,[10]

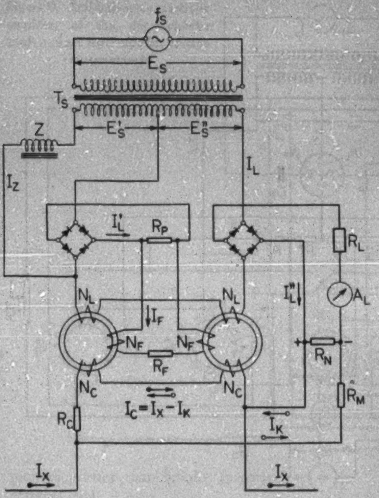
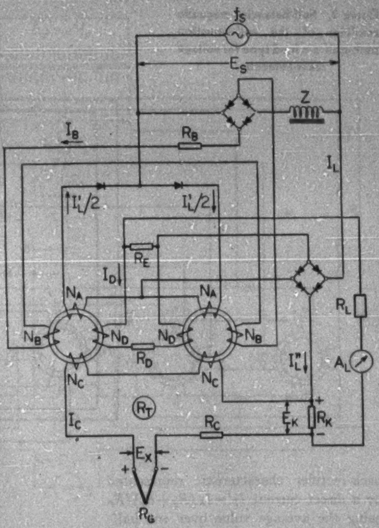

f_S
E_S
T_S
Z
E'_S
E"_S
I_Z
I_L
I'_L
R_P
R_L
N_L
N_L
I_F
I"_L
A_L
N_F
N_F
R_F
R_N
N_C
N_C
R_M
R_C
I_C=I_X−I_K
I_K
I_X
I_X

Figure 3 (left). Self-balancing magnetic amplifier of the external feedback type for current measurements. The current ratio of this circuit is $I_L/I_X=(R_N+R_M)/R_N$

Figure 5 (right). Self-balancing magnetic amplifier of the self-saturation type for voltage measurements. The transconductance of this circuit is $I_L/E_X=1/R_K$


f_S
E_S
Z
I_B
R_B
I_L
I'_L/2
I'_L/2
R_E
I_D
N_A
N_A
R_L
N_B
N_D
N_D
N_B
R_D
I"_L
A_L
N_C
N_C
R_T
I_C
R_C
E_K
R_K
E_X
R_G

total ampere-turns of load windings N_L are equal to total ampere turns of control windings N_C and feedback windings N_F

$$I_LN_L=I_CN_C+I_FN_F=\frac{E_X-I_LR_K}{R_T}N_C+I_LN_F\frac{R_P}{R_P+R_F} \quad (2)$$

When a voltage E_X is applied to the control circuit, the load current I_L increases so that a voltage drop $E_K=I_LR_K$ takes place across resistor R_K. Because the polarity of E_K opposes an increase in load current I_L, the resistor R_K exercises a controlling effect on the load current I_L.

Referring to equation 2 the relationship between E_X and I_L will be given by the equation

$$\frac{E_X}{I_L}=R_K+R_T\frac{N_L-N_F\,R_P/(R_P+R_F)}{N_C} \quad (3)$$

Thus using positive (regenerative) external feedback with $N_F\,R_P/(R_P+R_F)=N_L$ we get

$$\frac{E_X}{I_L}=R_K \quad \text{or} \quad \frac{I_L}{E_X}=\frac{1}{R_K} \quad (4)$$

where $1/R_K$ represents the "transconductance" of the circuit.

Referring to equation 4 it is evident that with $N_F\,R_P/(R_P+R_F)=N_L$ the voltage drop I_LR_K across resistor R_K will be equal to the voltage E_X to be measured, and the control current $I_C=(E_X-I_LR_K)/R_T$ will be zero.

This condition: $N_F\,R_P/(R_P+R_F)=N_L$ involves the assumption of an ideal feed-

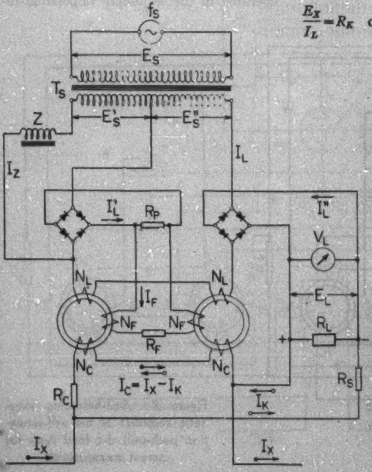

f_S
E_S
T_S
Z
E'_S
E"_S
I_Z
I_L
I'_L
R_P
I"_L
V_L
N_L
N_L
I_F
E_L
N_F
N_F
R_L
R_F
N_C
N_C
R_S
R_C
I_C=I_X−I_K
I_K
I_X
I_X

Figure 4 (left). Self-balancing magnetic amplifier of the external feedback type for current measurements. The transresistance of this circuit is $E_L/I_X=R_S$

Figure 6 (right). Self-balancing magnetic amplifier of the self-saturation type for current measurements. The current ratio of this circuit is $I_L/I_X=(R_N+R_M)/R_N$

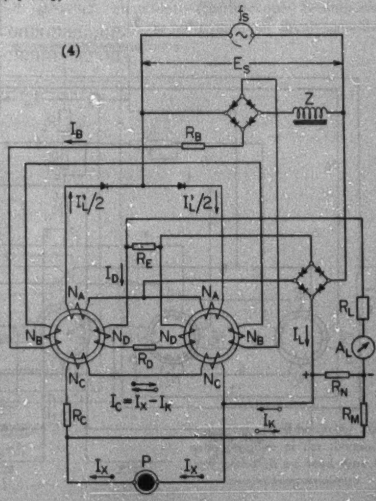
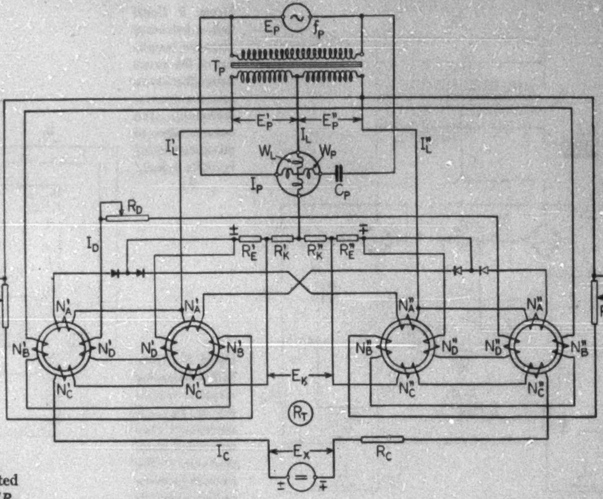

f_S
E_S
Z
I_B
R_B
I'_L/2
I'_L/2
R_E
I_D
N_A
N_A
R_L
N_B
N_D
N_D
N_B
I_L
A_L
R_D
R_N
N_C
N_C
I_C=I_X−I_K
R_C
I_K
R_M
I_X
P
I_X

Figure 7. Self-balancing magnetic amplifier of the self-saturation push-pull d-c load type for voltage measurements



back-rectifier characteristic represented by a direct current $I_L' = I_F(R_P + R_F)/R_P$ being the average value over one half-cycle of the alternating current I_L flowing through the load windings N_L. In practice, however, departure of the feedback rectifier from the ideal characteristic causes a certain reduction of current ratio I_L'/I_L so that actual feedback ampere-turns $I_F N_F$ also are reduced. By proper adjustment of the shunting resistor R_P acting as regeneration control, it is possible to compensate this effect of feedback-rectifier imperfections.

If the control windings N_C are temporarily made ineffective, for example, by disconnecting the input source R_G ($R_T = \infty$), and the regeneration control resistor R_P increased, a point of instability will be reached where the well-known trigger action due to excessive positive feedback occurs. At this point the circuit is critically regenerated and is essentially a magnetic amplifier of infinite gain. If the control windings N_C are then reconnected to the degenerative coupling resistor R_K, all tendency toward instability and trigger action will be restrained by superimposed degeneration, and the magnetic amplifier will operate in the usual way under extremely stable conditions.

It is interesting to note that the regeneration control resistor R_P may be further increased and the magnetic amplifier operated in the seemingly impossible re-

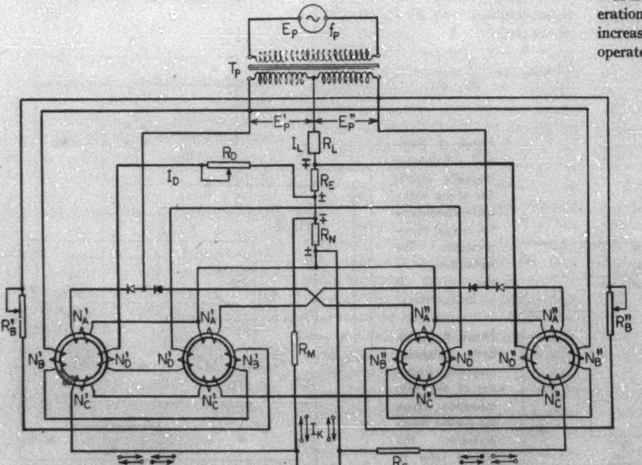
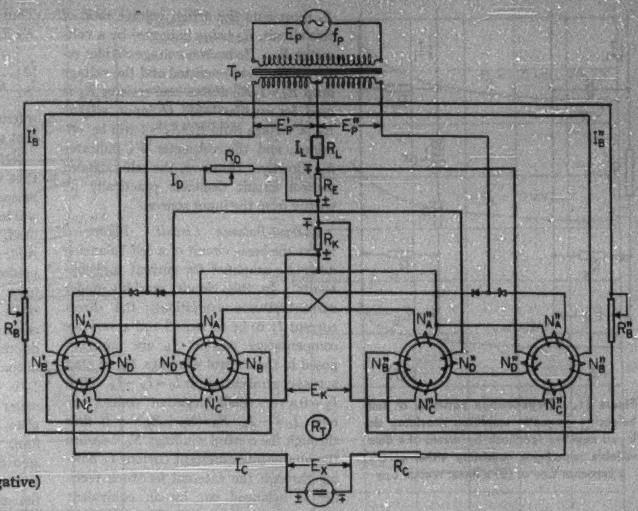


Figure 8. Self-balancing magnetic amplifier of the self-saturation push-pull d-c load type for current measurements

Figure 9. Self-balancing magnetic amplifier of the ringmodulator push-pull a-c load type for voltage measurements



gion of greater than infinite (negative) internal gain.

With normal working conditions corresponding to adjustment of R_F to the point of critical regeneration the control current $I_C=(E_X-I_LR_X)/R_T$ is zero, and the control windings N_C exert predominantly a transient type of control upon the two saturable reactor elements, while the regenerative external feedback windings N_F and the a-c load windings N_L do the actual work of establishing the output current I_L. If the control current I_C is observed by means of a microammeter while an input voltage E_X is applied, no change is noticeable. Therefore, this voltage-balance condition of zero control current I_C may be used in practice as the criteria for adjustment of the regeneration control resistor R_F: If R_F is adjusted below or above the point of critical regeneration while the control circuit is subjected to variations of input voltage E_X, a control current I_C will appear but of a direction of current flow dependent upon the direction of actual misadjustment.

For practical applications of a magnetic amplifier, as shown in Figure 1, the unique feature of practically zero control current I_C for the entire working range of output current I_L means a materially more linear relationship between voltage E_X and current I_L represented by equation 4. The self-balanced control circuit demands practically no energy from the input source (thermocouple) because even with the optimum input circuit matching conditions

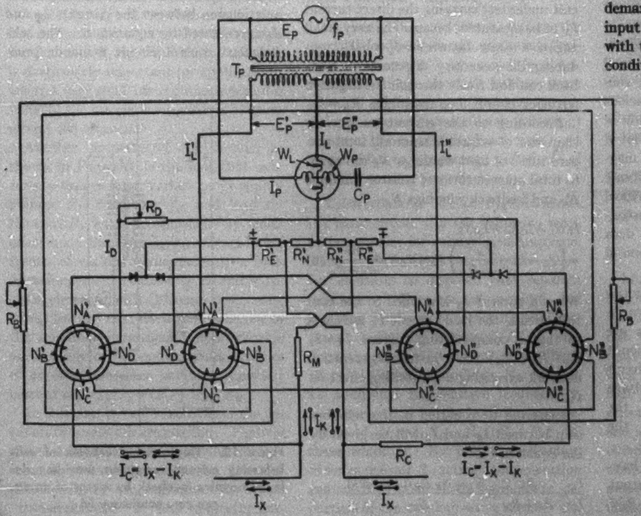


Figure 10. Self-balancing magnetic amplifier of the ringmodulator push-pull a-c load type for current measurements

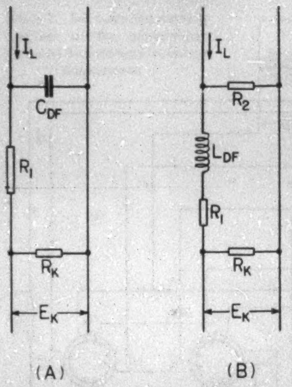


Figure 11. Degeneration networks of self-balancing magnetic amplifiers providing delayed negative feedback by means of a time-variable impedance consisting either of (A) a capacitor C_{DF} or (B) a linear reactor L_{DF}

$$R_C+R_X=R_G=R_T/2 \qquad (5)$$

the effective input resistance E_X/I_C is theoretically infinite and is practically about 100 to 1,000 times higher than R_C+R_X. Of course, this fact is particularly valuable for applications, where a very high input circuit resistance value of the magnetic amplifier is of paramount importance.

Another valuable property of the circuit shown in Figure 1 is the fact that over certain ranges large variations in supply voltage E_S, frequency f_S, and load resistance R_L have a comparatively small effect upon the ratio E_X/I_L. Indeed, consideration of the results of various measurements on such a circuit leads to the conclusion that the actual behavior of a self-balancing potentiometer type magnetic amplifier approximates very closely corresponding characteristics of an "ideal amplifier" having a constant transconductance $I_L/E_X=1/R_K$, just as does an electronic amplifier of the compound feedback type.[16]

In accordance with common practice in the field of d-c amplifiers, some variations of the degeneration network are possible. For simplicity of explanation it has been assumed that the basic circuit of Figure 1 is arranged to provide voltage input E_X with current output I_L, the value $I_L/E_X=1/R_K$ representing the transconductance of this circuit.

Figure 2 shows a modified circuit having voltage input E_X and voltage output E_L. The expression $E_L/E_X=(R_1+R_2)/R_2$ represents the actual voltage ratio of this circuit, E_L being indicated by a voltmeter V_L. If the two voltage divider resistors R_1, R_2 are omitted and the voltage drop $E_L=I_LR_L$ across load resistor R_L is used for compensation of input voltage E_X, a voltage ratio $E_L/E_X=1$ will be obtained, and the voltmeter V_L indicates E_X directly. However, the self-balanced control circuit demands practically no energy from the input source.

Current-Balance Circuit. Figure 3 shows the basic circuit of a self-balancing magnetic amplifier for current measurements. In this circuit, working under current-balance conditions, the direct current I_X to be measured and a variable compensating current I_K are superimposed in the control windings N_C, so that actual control current is $I_C=I_X-I_K$. With $I_X=0$ a very small "negative" control current $I_C=-I_L\ R_N/(R_N+R_M)$ will flow through the control windings N_C, because the alternating quiescent current I_L flowing through the external feedback rectifier is balanced out by an equivalent alternating current I_Z. The balancing current I_Z, exercising a bias effect, is produced by a center-tapped supply transformer T_S and a nickel-iron-alloy-type series impedance Z. Of course, this a-c bias circuit may be provided in a voltage-balance arrangement, as shown in Figure 1, instead of a bias circuit based upon the use of d-c bias ampere-turns I_BN_B.

Note the current-balance arrangement shown in Figure 3 does not need an external conduction path through the circuit under test carrying the direct current I_X to be measured, because the very small negative control current $I_C=-I_K$ producing the necessary degenerative feedback can flow freely through the negative feedback circuit loop resistance R_M+R_C.

According to the current-transformer character of saturable reactors, total ampere-turns of load windings N_L are equal to total ampere-turns of control windings N_C and feedback windings N_F

$$I_LN_L=I_CN_C+I_FN_F$$
$$=\left(I_X-I_L\frac{R_N}{R_N+R_M}\right)N_C+I_LN_F\frac{R_P}{R_P+R_F} \qquad (6)$$

When a current I_X is applied to the control circuit, the load current I_L increases so that a compensating current $I_K=I_L\ R_N/(R_N+R_M)$ flows in the degeneration network exercising a controlling effect on I_L.

Referring to equation 6, the relationship between I_X and I_L will be given by the equation

$$\frac{I_X}{I_L}=\frac{R_N}{R_N+R_M}+\frac{N_L-N_F\,R_P/(R_P+R_F)}{N_C} \qquad (7)$$

Thus using "critical regeneration" with $N_F\,R_P/(R_P+R_F)=N_L$ we get

$$\frac{I_X}{I_L}=\frac{R_N}{R_N+R_M} \quad \text{or} \quad \frac{I_L}{I_X}=\frac{R_N+R_M}{R_N} \qquad (8)$$

where $(R_N+R_M)/R_N$ represents the "current ratio" of the circuit.

Referring to equation 8, it is evident that with critical regeneration, the compensating current $I_K=I_L\ R_N/(R_N+R_M)$ will be equal to the current I_X to be measured, control current $I_C=I_X-I_K$ will be zero, and the control windings N_C exert predominantly a transient type of control upon the saturable reactor elements, while the regenerative external feedback windings N_F and the a-c load windings N_L do the actual work of establishing the output current I_L. If the control current I_C is observed by means of a microammeter while an input current I_X is applied, no change is noticeable. Therefore, this current-balance condition of zero control current I_C may be used in practice as the criteria for adjustment of the regeneration control resistor R_P: If R_P is adjusted below or above the point of critical regeneration while the control circuit is subjected to variations of input current I_X, a control current I_C will appear, but of a direction of current flow dependent upon the direction of actual misadjustment.

For practical applications of a magnetic amplifier, as shown in Figure 3, the unique feature of practically zero control current I_C for the entire working range of output current I_L means a materially more linear relationship between the currents I_X and I_L represented by equation 8. The self-balanced control circuit demands prac-

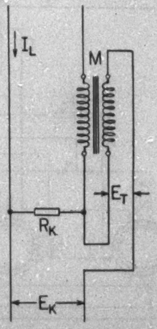


Figure 12. Degeneration network of self-balancing magnetic amplifiers providing delayed negative feedback by means of an air-gap core transformer M

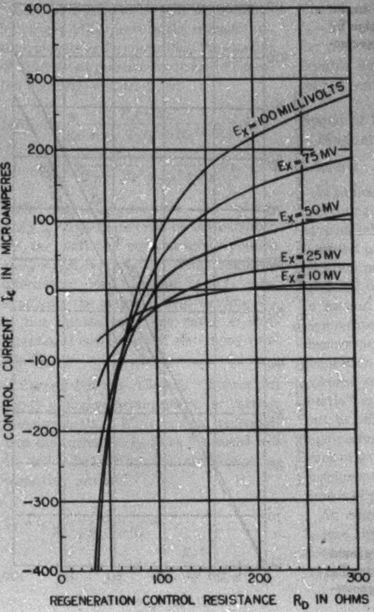
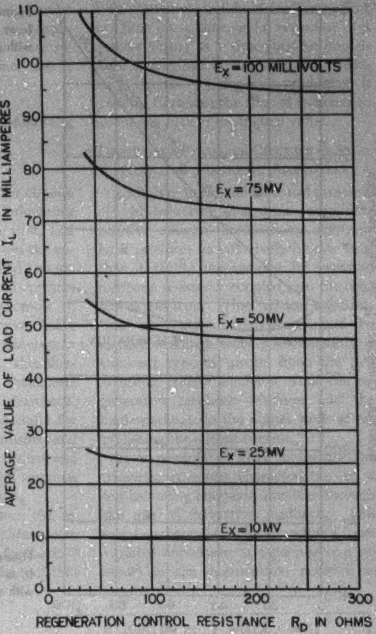


Figure 13 (left). Control current I_C as a function of the regeneration control resistor R_D with voltage E_X as a parameter

Figure 14 (right). Load current I_L as a function of the regeneration control resistor R_D with voltage E_X as a parameter



tically no energy from the circuit under test carrying the direct current I_X because, even with optimum input circuit matching conditions, the effective input voltage $(I_X-I_K)\ R_C$ is theoretically zero and is practically about 100 to 1,000 times smaller than $I_X R_C$. Evidently, this fact is particularly important for applications, where a very small input circuit resistance value of the magnetic amplifier is of paramount importance.

Another property of the circuit, as shown in Figure 3, is the fact that over certain ranges large variations in supply voltage E_S, frequency f_S, and load resistance R_L have a comparatively small effect upon the ratio I_X/I_L. Considering results of various measurements on such an arrangement containing saturable reactor elements with Orthonol tape cores, it was learned that the actual behavior of the amplifier investigated approximates very closely corresponding characteristics of an ideal amplifier having a constant current ratio $I_L/I_X=(R_N+R_M)/R_M$.

In this circuit, too, variations of the degenerating network are possible. Figure 4 shows a modified circuit having current input I_X and voltage output E_L. The expression $E_L/I_X=R_S$ represents the actual "transresistance" of this magnetic amplifier, E_L being indicated by a voltmeter V_L.

SELF-FEEDBACK CIRCUITS

Obviously it is possible to replace the external feedback circuit producing "critical regeneration" by a self-feedback (self-saturation) circuit[18] having some practical advantages. However, an additional regenerative feedback (boosting) circuit of the external feedback type must be provided to compensate the effects of dry-disk rectifier imperfections.

Voltage-Balance Circuit. Figure 5 shows the basic circuit of a self-balancing magnetic amplifier for voltage measurements. This circuit contains two half-wave rectifier elements so disposed that separate feedback windings are not required, and the load windings N_A carrying unidirectional (pulsating) currents $(I_L'/2)$ produce the desired regenerative feedback effects. Boosting windings N_D exercise an additional regenerative feedback effect represented by the boosting ampereturns $I_D N_D=I_L N_D \quad R_S/(R_S+R_D)$ of these windings. By proper adjustment of series resistor R_D acting as regeneration control it is possible to compensate departure of the two half-wave rectifier elements from the ideal characteristics and to obtain critical regeneration.

Current-Balance Circuit. Figure 6 shows the basic circuit of a self-balancing magnetic amplifier for current measurements. This arrangement, representing a modification of the current-balance circuit shown in Figure 3, uses self-saturation and boosting windings N_D to provide critical regeneration by proper adjustment of the regeneration control resistor R_D.

Push-Pull Magnetic Amplifier Circuits of the Self-Balancing Potentiometer Type

GENERAL

Magnetic amplifiers of the balance detector type[8] are based upon the use of symmetrical push-pull circuits which contain two balanced saturable reactor systems utilizing positive and negative feedback effects. The resultant output current of these amplifiers, either direct or alternating current, is duo-directional and is zero at zero input. Such push-pull circuits are especially useful for applications in the fields of electrical measurements, high-performance servomechanisms, and automatic control.

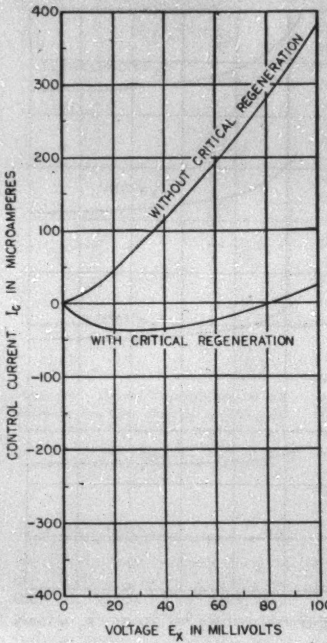


Figure 15 (left). Control current I_C as a function of the voltage E_X with and without critical regeneration

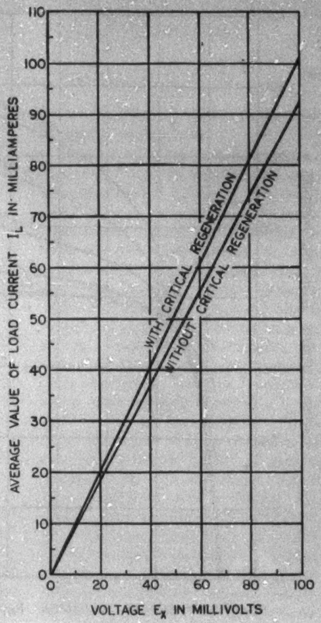


Figure 16 (right). Load current I_L as a function of the voltage E_X with and without critical regeneration

It is possible to apply the principle of self-balancing compound feedback circuits, see Figures 1 to 6, using critical regeneration for the design of push-pull magnetic amplifier circuits of the balance detector type. Four typical examples of such circuits will be described in the following paragraphs.

SELF-SATURATING PUSH-PULL CIRCUITS WITH D-C LOAD

Voltage-Balance Circuit. Figure 7 shows a push-pull-type self-balancing magnetic amplifier for voltage measurements. This arrangement contains two balanced saturable reactor systems utilizing compound feedback and voltage-balance conditions in the input circuit. A certain amount of regenerative feedback is provided by the load windings N_A', N_A'' and the associated half-wave dry-disk rectifier elements producing self-saturation. Using properly rated a-c bias circuits N_B', R_B' and N_B'', R_B'' carrying pure alternating bias currents I_B', I_B'', it will be possible to provide the desired quiescent current value of the two direct output currents I_L' and I_L'' giving the resultant direct load current $I_L=I_L'-I_L''$ which flows through the d-c load resistor R_L, for example a moving-coil-type ink recorder.

Critical regeneration will be obtained by proper adjustment of the regeneration control resistor R_D controlling the amount of differential positive feedback (differential boosting) ampere-turns

$$I_DN_D=I_L\frac{R_E}{R_E+R_D}N_D \tag{9}$$

With critical regeneration the voltage $\pm E_X$ to be measured will be self-balanced by the variable compensating voltage $\pm E_K=\pm I_LR_K$. The direct load current $I_L=I_L'-I_L''$ supplying the d-c load resistor R_L is duo-directional and is zero with zero input voltage E_X. The control circuit (total resistance R_T) operating under voltage balance conditions ($\pm E_K=\pm E_X$) demands practically no energy from the input source, even with optimum matching of the internal and external resistance of the input circuit loop.

Current-Balance Circuit. Figure 8 shows a similar push-pull-type self-balancing magnetic amplifier for current measurements. With critical regeneration obtained by proper adjustment of the regeneration control resistor R_D the current I_X to be measured will be self-balanced by the variable compensating current $I_K=I_LR_N/(R_N+R_M)$. The resultant direct load current $I_L=I_L'-I_L''$ flowing through the d-c load resistor R_L, for example an ink recorder, is duo-directional and is zero at zero input current I_X. The control circuit operating under current-balance conditions demands practically no energy from the circuit under test carrying the input current I_X, even when optimum matching is used.

RINGMODULATOR TYPE SELF-SATURATING PUSH-PULL CIRCUITS WITH A-C LOAD

Voltage-Balance Circuit. The arrangement shown in Figure 9 contains two balanced saturable reactor systems utilizing compound feedback and voltage-balance conditions in the input circuit. A certain amount of regenerative feedback is produced by load windings N_A', N_A'' co-operating with four half-wave dry-disk rectifier elements in a ringmodulator type self-saturating push-pull circuit. Using properly rated a-c bias circuits N_B', R_B' and N_B'', R_B'' carrying pure alternating bias currents I_B', I_B'' it will be possible to provide the desired quiescent current value of the two alternating output currents I_L' and I_L'' which flows through the amplifier field winding W_L of a 2-phase induction-type reversible motor having separately excited line field winding W_P with a series capacitor C_P.

Critical regeneration will be obtained by proper adjustment of the regeneration control resistor R_D controlling the amount of differential positive feedback (differential boosting) ampere-turns

$$I_D N_D = I_L \frac{R_B'+R_K'}{R_B'+R_K'+R_B''+R_K''+R_D} N_D = \frac{R_B''+R_K''}{R_B'+R_K'+R_B''+R_K''+R_D} N_D \quad (10)$$

With critical regeneration the voltage $\pm E_X$ to be measured will be self-balanced by the variable compensating voltage $\pm E_K = I_L R_K' = I_L R_K''$. The resulting alternating load current $I_L = I_L' - I_L''$ supplying the amplifier field winding W_L of the 2-phase reversible motor is duo-directional and is zero at zero input voltage E_X.

Current-Balance Circuit. Figure 10 shows a similar arrangement for current measurements. With critical regeneration the current $\pm I_X$ to be measured will be self-balanced by the variable compensating current

$$I_K = I_L \frac{R_N'}{R_N'+R_N''+R_M} = I_L \frac{R_N''}{R_N'+R_N''+R_M} \quad (11)$$

The resulting alternating load current $I_L = I_L' - I_L''$ flowing through the amplifier field winding W_L of the 2-phase reversible motor is duo-directional and is zero at zero input current I_X.

Speed of Response in Self-Balancing Magnetic Amplifiers

GENERAL

In principle, magnetic amplifier circuits containing a large amount of inductance have an inherent slow response time, and there is a certain time lag between the application of the input signal and the attainment of the full output load current. In certain applications, for example in high-temperature measurements using thermocouples or radiation pyrometers in connection with ink recorders, this sluggishness of response due to "magnetic inertia" is not an important drawback. But in other applications, especially for high-performance servomechanisms and some other types of automatic control and instrumentation, a very short response time is of paramount importance.

An efficient method to reduce the response time of magnetic amplifiers consists in providing "derivative feedback."[8] This method is based upon the fact that here actual magnitude of feedback effect is a function of time, and a capacitor or a reactor acting as a time-variable impedance is provided in the "derivative feedback circuit." Using this method it is possible to reduce the time of response to about one-fifth of its original value.

DERIVATIVE FEEDBACK CIRCUITS OF SELF-BALANCING MAGNETIC AMPLIFIERS

According to the fundamental principle of self-balancing magnetic amplifiers positive external feedback or self-feedback produces an effectively infinite gain that is highly degenerated by negative feedback resistance-coupled into the control circuit loop. With "critical regeneration" the control current is zero, and the control windings exert predominantly a transient type of control upon the two saturable reactor elements, while the regenerative feedback windings and the load windings do the actual work of establishing the output current.

It is possible to alter materially the transient response characteristics of a self-balancing magnetic amplifier through the use of derivative feedback. This may be accomplished by providing a time-variable impedance (a capacitor or a reactor) in the degeneration network in

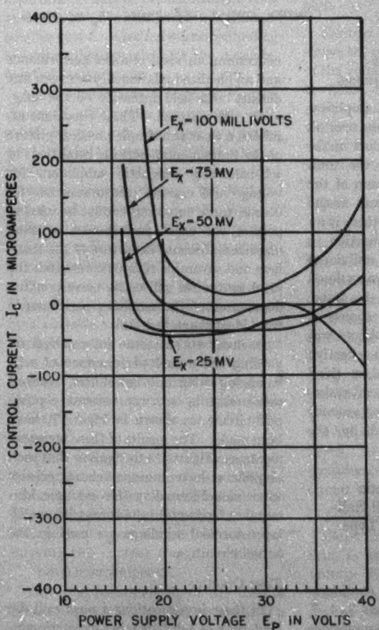
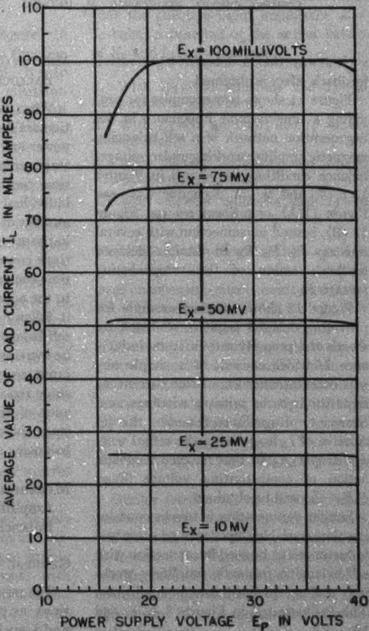


Figure 17 (left). Control current I_C as a function of power supply voltage E_P with voltage E_X as a parameter

Figure 18 (right). Load current I_L as a function of power supply voltage E_P with voltage E_X as a parameter



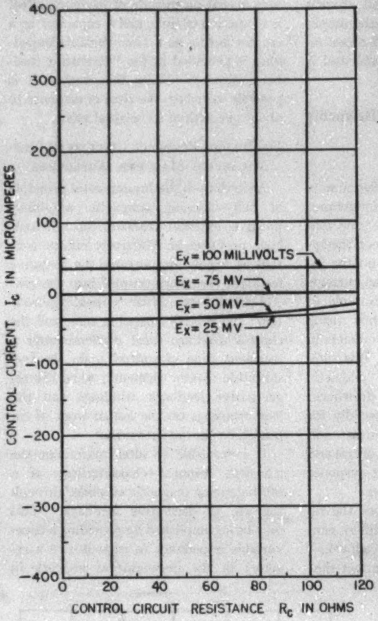


Figure 19 (left). Control current I_C as a function of the control circuit resistance R_C with voltage E_X as a parameter

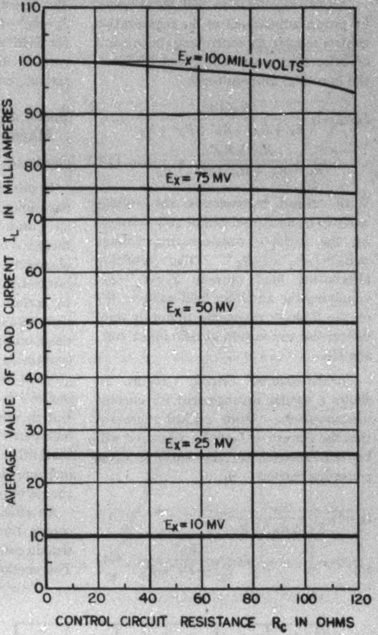


Figure 20 (right). Load current I_L as a function of the control circuit resistance R_C with voltage E_X as a parameter

such a way that a delayed degenerative feedback effect is obtained.

Figure 11 shows two examples for providing a time-variable impedance in the degeneration network of a self-balancing magnetic amplifier working under voltage-balance conditions, as shown in Figures 1, 5, 7, and 9. A capacitor C_{DF}, see Figure 11(A), or a linear reactor, Figure 11 (B), is used in connection with several resistors R_1, R_2, R_K to obtain a delayed feedback action of the compensating voltage E_K.

Figure 12 shows another example for providing delayed negative feedback by means of a properly ratedm utual inductance M which consists of a simple air-gap core transformer: Load current I_L flows through the primary windings, and transient voltage E_T representing the derivative of I_L is opposing the actual voltage drop I_LR_K so that delayed feedback action of compensating voltage $E_K = I_LR_K - E_T$ will be obtained.

Similar degeneration networks containing time-variable impedance or a mutual inductance can be used in connection with self-balancing magnetic amplifiers working under current-balance conditions. This is illustrated in Figures 3, 6, 8, and 10.

MULTISTAGE CIRCUITS OF SELF-BALANCING MAGNETIC AMPLIFIERS

When two or three stages of amplification are connected in cascade, the over-all power amplification is the product of the stage amplifications, whereas the total time constant is roughly the sum of the individual time constants. Thus, as the number of stages is increased, there is an increase of the ratio of amplification to time constant. Because the individual time constant of each stage is proportional to the actual load resistance of this stage, it follows that the speed of response of self-balancing multistage amplifiers will be increased by the fact that the effective input resistance of the second or third stage representing the actual load resistance of the preceding (first or second) stage is considerably reduced by the balancing effects.

Measurements on a Magnetic Amplifier of the Push-Pull Self-Balancing Potentiometer Type

GENERAL

The results of the theoretical considerations, as presented by equations 3, 4, 7, and 8, are based upon certain assumptions concerning an ideal rectifier performance and an idealized relationship between flux density and field intensity of the magnetic core material. These equations establish a clear and simple physical picture of the fundamental working conditions of self-balancing magnetic amplifiers for voltage and current measurements. Of course, these equations cannot be used to predict the exact performance of such amplifiers, because departure of the rectifiers and saturable reactor cores from the ideal properties will cause corresponding departures of the amplifier performance from ideal characteristics.

In order to get some information regarding the actual performance of self-balancing magnetic amplifiers, various measurements on a representative push-pull circuit, as shown in Figure 7, have been made. The results of these measurements, see Figures 13 to 20, show that such amplifiers have unique characteristics when nickel-iron-alloy (for example Mumetal or Orthanol) tape cores with multilayer toroidal windings are used in the actual circuit.

TEST CIRCUIT

In these investigations a push-pull d-c load type of self-balancing magnetic-amplifier circuit for voltage measurements, see Figure 7, has been used. This circuit contains four Mumetal 2-mil tape cores (outside diameter, 2 inches; inside diameter, $1^5/_8$ inches; tape width, 0.5 inch) having multilayer toroidal windings ($N_A=2{,}000$, $N_B=50$, $N_C=500$, $N_D=200$, each with number 28 wire size). The magnetic circuit has an effective area of 0.084 inch square = 0.54 centimeter square and a mean length of 5.7 inches = 14.5 centimeters.

Selenium dry-disk rectifier elements are used in such a way that the maximum value of reverse voltage on each rectifier element does not exceed about 10 volts, and favorable working conditions are obtained regarding the reverse resistances and the corresponding reverse-current components of the balanced self-saturating circuits.

The d-c load resistor $R_L=5$ ohms and the direct load current $I_L=-100\ldots0\ldots+100$ milliamperes are rated to give a maximum output power of 50 milliwatts, corresponding to the power requirements of a conventional moving-coil-type ink recorder. In this example, it is the purpose of the self-balancing magnetic amplifier to extend the sensitivity of the ink recorder so that it will be possible to make a permanent record of small direct voltages created, for instance, by thermocouples or d-c bridge networks containing resistance thermometers for the measurement of temperatures.

Actual power supply voltage E_P (frequency $f_P=60$ cycles per second) indicated by an iron-vane-type voltmeter can be varied by means of a Variac transformer. The actual value of direct load current I_L flowing through the load resistor (ink recorder) $R_L=5$ ohms is measured by means of a precision moving-coil-type milliammeter (current range: $0\ldots100$ milliamperes).

The direct voltage E_X to be measured ($E_X=0\ldots100$ millivolts) is represented by a variable voltage drop across a fixed 10-ohm resistor, and the variable current from a storage battery which flows through this resistor is indicated by a second precision moving-coil-type milliammeter (current range: $0\ldots10$ milliamperes).

Actual value of control current I_C flowing through the control circuit loop is measured by means of a high-sensitivity galvanometer which indicates the actual voltage drop across a fixed 1-ohm resistor carrying this current (range: $-400\ldots0\ldots+400$ microamperes).

The total resistance R_T of the control circuit loop can be changed by means of a variable series resistor R_C, and the differential feedback current I_D can be changed by means of a variable series resistor R_D, to provide critical regeneration, as defined by equations 3 and 7.

The following results of various measurements show the relationship between the involved six parameters: E_X, E_P, I_L, I_C, R_D, and R_C.

RESULTS OF MEASUREMENTS AND DISCUSSION

Adjustment of the Regeneration Control Resistor. Figure 13 shows the control current I_C as a function of the regeneration control resistor R_D with voltage E_X as a parameter. It is evident that the optimum value of R_D giving full balance conditions ($I_C=0$) depends upon the actual magnitude of E_X, and that $R_D=80$ ohms will cause a control current $I_C\approx40$ microamperes to flow through the control circuit loop when E_X is varied within the limits from 0 to 100 millivolts.

Figure 14 shows the average value of load current I_L as a function of the regeneration control resistor R_D with voltage E_X as a parameter. These characteristic curves illustrate the importance of proper adjustment of R_D to get a substantially constant value of the transconductance $I_L/E_X\approx1/R_K$ with $R_K=1.0$ ohm.

Transfer Characteristics. Figure 15 shows the control current I_C as a function of the voltage E_X to be measured. It is to be noted that $I_C=0\ldots+380$ microamperes when no critical regeneration circuit will be used ($R_D=\infty$), and that I_C varies within the limits from $-40\ldots0\ldots+25$ microamperes with proper adjustment of the regeneration control resistor, see Figure 13; $R_D=80$ ohms.

Figure 16 shows the average value of load current I_L as a function of E_X with and without critical regeneration. Note that there is a linear relationship between output current I_L and input voltage E_X in both cases. However, there is a considerable departure of transconductance I_L/E_X from its ideal value (about 10 per cent), if no critical regeneration has been provided. The load current values are too small in the entire working range of the amplifier, because the control current $I_C=0\ldots+380$ microamperes causes a considerable voltage drop in the control circuit loop.

Influence of Changes in Power Supply Voltage. Figure 17 shows the control current I_C as a function of power supply voltage $E_P=10\ldots40$ volts (rms) with voltage E_X as a parameter. It is interesting to note that $I_C=-40\ldots0\ldots+40$ microamperes when E_P will be changed within comparatively wide limits ($22\ldots28\ldots34$ volts) corresponding to variations of about ±20 per cent.

Figure 18 shows the average value of load current I_L as a function of E_P. It is evident that the transconductance I_L/E_X remains constant when E_P will be changed within the aforementioned limits.

Influence of Changes in Control Circuit Resistance. Figure 19 shows the control current I_C as a function of the control circuit resistance R_C with voltage E_X as a parameter. The rated value of total control circuit resistance is $R_T=R_G+R_N+R_C=10+16+14=40$ ohms. It is to be noted that R_C has been changed within excessively wide limits ($0\ldots120$ ohms) to illustrate the influence of such very large changes in control circuit resistance which, will never occur in a conventional type of thermocouple input circuit.

Figure 20 shows the average value of load current I_L as a function of the control circuit resistance R_C with E_X as a parameter. These characteristic curves illustrate the fact that, even with extremely large changes of R_C, the transconductance I_L/E_X remains substantially constant.

Effective Input Impedance. Referring to Figures 13, 15, 17, and 19, it is evident that the effective input impedance E_X/I_C being a function of the actual values of E_X and I_C is, for example:

1. With $E_X=100$ millivolts and $I_C=40$ microamperes, 2,500 ohms.
2. With $E_X=75$ millivolts and $I_C=0$, extremely high.
3. With $E_X=50$ millivolts and $I_C=40$ microamperes, 1,250 ohms.
4. With $E_X=25$ millivolts and $I_C=40$ microamperes, about 600 ohms.

These relationships explain the aforementioned fact that, even with extremely large changes in control circuit resistance, the transconductance I_L/E_X will have a substantially constant value.

Of course, it is possible to adjust the regeneration control resistor R_D in such a way that the balance conditions will be improved in the range of small voltages $E_X=0\ldots50$ millivolts, the input impedance remaining sufficiently high with higher values of E_X ($50\ldots100$ millivolts).

Similar results will be obtained, when current-balance conditions, see Figure 8, will be used. The effective voltage drop R_TI_C across the terminals of the control windings N_C (total resistance R_T) will be correspondingly reduced by the fact that the current I_X to be measured is substantially balanced out by the variable compensating current $I_K = I_L R_N/(R_N + R_M)$ so so that only a very small control current $I_C = I_X - I_K$ will flow through the control windings N_C.

Applications of Magnetic Amplifiers of the Self-Balancing Potentiometer Type

General

In order to describe practical applications of these magnetic amplifiers in the fields of electrical measurements, telemetering, and automatic control, and for the operation of small d-c or a-c reversing motors working in connection with high-performance servomechanisms, it is important to stress the following facts:

1. Self-balancing magnetic amplifiers operating under voltage-balance or current-balance conditions of the control circuit demand practically no energy from the d-c input source producing the voltage or current to be measured.

2. The effective input resistance of a self-balancing magnetic amplifier for voltage measurements is practically about 100 to 1,000 times higher than the actual input circuit loop resistance.

3. The effective input resistance of a self-balancing magnetic amplifier for current measurements is practically about 100 to 1,000 times smaller than the actual resistance of the control windings.

4. The unique feature of a considerably reduced control current for the entire working range of the self-balancing magnetic amplifier means a materially more linear relationship between the input and output quantities.

5. Over certain ranges, large variations in supply voltage, supply frequency, load resistance, and control circuit resistance have a comparatively small effect upon the transconductance, voltage ratio, current ratio, or transresistance of self-balancing magnetic amplifiers operating under extremely stable conditions with a high overall power gain.

6. For certain applications special compensating effects can be applied by designing the degeneration network in such a way that resistance values of this network are a function of voltage, current, temperature, or time (derivative feedback effects, for example, delayed degenerative feedback).

7. Results of various measurements, see the section entitled "Measurements on a Magnetic Amplifier of the Push-Pull Self-Balancing Potentiometer Type," lead to the conclusion that the actual behavior of self-balancing magnetic amplifiers approximates very closely corresponding characteristics of an ideal amplifier, in a way similar to an electronic amplifier of the compound feedback type.[16]

Measurement of Electrical Quantities

Magnetic amplifiers of the self-balancing potentiometer type are particularly suitable for extending the sensitivity of direct indicating electrical (multirange) instruments and ink recorders making a permanent record of very small direct voltages and currents, as created, for instance by thermocouples, photoelectric cells or radiation pyrometers. These high-performance d-c amplifiers also can be used in connection with phase-sensitive rectifiers for the measurement of various a-c quantities, for example, in a-c potentiometer circuits and a-c bridge networks. Voltage dividers and shunting arrangements may be provided to extend the voltage and current ranges of such amplifiers.

Measurement of Nonelectrical Quantities

In accordance with common practice nonelectrical quantities will be measured by means of transducers providing the desired transition in an electrical value. There are numerous kinds of transducers representing variable resistors, inductances, mutual inductances, capacitors, and so forth.[17]

Special Applications of Self-Balancing Magnetic Amplifiers

Referring to Figure 1, it is evident that the self-balancing magnetic amplifier demands practically no energy from the thermocouple, because the effective input impedance (E_X/I_C) is practically about 100 to 1,000 times higher than the actual resistance of the control circuit loop. Therefore, comparatively large changes in resistance of the thermocouple or other resistors contained in the control circuit will have no appreciable effect upon the accuracy of measurement.

Photoelectric cells of the selenium barrier layer dry-disk type find wide application in the measurement and control of many industrial processes, including both mechanical and chemical. When these cells are used in circuits of low external resistance, the variation of the photoelectric current with respect to the illumination is practically linear. Referring to Figure 6, it is evident that the self-balancing magnetic amplifier demands no energy from the photocell P, because, even with optimum input circuit matching conditions, the effective external resistance is practically about 100 to 1,000 times smaller than the actual resistance value of the control windings of the amplifier. Therefore, the voltage across the photoelectric cell will be practically zero, and the relationship between current and illumination is then substantially linear.

Although such photoelectric cells generally develop ample current for direct operation of moving-coil-type instruments and sensitive relays without the use of amplifiers, it will be necessary in special cases to provide a simple and reliable d-c amplifier, for example, for operation of moving-coil-type ink recorders. Evidently, it is possible to extend the sensitivity of this type of ink recorders. This will be useful in many other cases, especially for the purpose of telemetering.

Power measurements can be made at high frequencies by dissipating the power in a bolometer, which is a resistive element with a large temperature coefficient of resistance. The magnitude of this power can be determined from the measured change in resistance of a bolometer bridge operating in connection with a self-balancing magnetic amplifier and a moving-coil-type ink recorder.

References

Comprehensive bibliographies concerning magnetic amplifiers and their practical applications have been given in the following papers:

1. The Transductor (Sometimes Called "Saturable Reactor" or "Magnetic Amplifier"), H. B. Rex. *Instruments* (Pittsburgh, Pa.), volume 20, December 1947, pages 1102–09, volume 21, April 1948, pages 332, 352–62.

2. Bibliography of Magnetic Amplifier Devices and the Saturable Reactor Art, James G. Miles. *AIEE Transactions*, volume 70, part II, 1951, pages 2104–17.

Application of saturable reactors for the measurement of large direct currents and high direct voltages are described in the following papers (reference 3: null method; references 4–7: deflection method):

3. The Transformer for Measuring Large Direct Currents, Harris J. Ryan. *AIEE Transactions*, volume 18, 1901, pages 169–83.

4. Messung Starker Gleichströme auf Grosse Entfernungen, E. Besag. *Elektrotechnische Zeitschrift* (Berlin, Germany), volume 40, 1919, pages 436–37.

5. Messung von Gleichströmen durch Vormagnetisierung von Wechselstrom-Drosseln, G. Keinath. *Archiv fur Technisches Messen* (Munich, Germany), Section V 3210-1, March 1933.

6. Ein Einfacher Gleichstrom-Messwandler mit Echten Stromwandler-Eigenschaften, W. Krämer. *Elektrotechnische Zeitschrift* (Berlin, Germany), volume 58, 1937, pages 1309–13.

7. Ein Neuer Gleichspannungs-Messwandler zur Messung Hoher Gleichspannungen, W. Krämer. *Elektrotechnische Zeitschrift* (Berlin, Germany), volume 59, 1938, pages 1295–1298.

8. Magnetic Amplifiers of the Balance Detector Type—Their Basic Principles, Characteristics, and Applicatons, W. A. Geyger. *AIEE Transactions*, volume 70, part II, 1951, pages 1707–20.

9. Magnetic Amplifiers with Orthonol Tape Cores, W. A. Geyger. *Proceedings*, National Electronics Conference (Chicago, Ill.), volume 6, 1950, pages 59–67.

10. Special Magnetic Alloys and Applications, G. W. Elmen, E. A. Gaugler. *Electrical Engineering*, volume 67, September 1948, pages 843–45.

11. Ein Verfahren zum Aufzeichnen Kleiner Gleichspannungen mit Tintenschrift, W. Geyger. *Archiv der Elektrischen Übertragung* (Wiesbaden, Germany), volume 3, 1949, pages 165–78.

12. Magnetic Amplifier, A. S. FitzGerald.

United States Patent Number *2,464,639*, March 15, 1949. (Figure 1).

13. Magnetic Amplifier System, A. S. FitzGerald. United States Patent Number *2,461,046*, February 8, 1949 (Figure 8).

14. Electromagnetic System, A. S. FitzGerald. United States Patent Number *2,459,069*, January 11, 1949 (Figures 2, 3, and 5).

15. Apparatus for Measuring Direct Currents or Voltages, W. Geyger. United States Patent Number *2,338,423*, January 4, 1944. (German Priority: December 14, 1938).

16. Stable Voltmeter, R. W. Gilbert. *Electronics*, McGraw-Hill Publishing Company (New York, N. Y.), March 1947, pages 130–33.

17. Transducers, J. R. Stovall. *Electrical Manufacturing* (New York, N. Y.), volume 45, April 1950, pages 88–92, 176–84.

18. Self-Saturation in Magnetic Amplifiers, W. J. Dornhoefer. *AIEE Transactions*, volume 68, part II, 1949, pages 835–50.

No Discussion

Ferroelectric Materials as Storage Elements for Digital Computers and Switching Systems

J. R. ANDERSON
NONMEMBER AIEE

THE growing need for improved types of data storage systems in both the digital computer and telephone switching fields has led to the investigation of ferroelectric materials as memory devices. Among the objectives in developing these devices have been low power consumption while storing or reading out information, memory access times of 1 microsecond or less, small size per bit of memory, simplified low cost construction, and high reliability. This paper presents an explanation of the operation of a basic storage circuit employing a ferroelectric material, requirements for suitable materials, descriptions of several types of circuits employing ferroelectrics, and experimental data obtained with operating circuits.

Ferroelectric Materials for Memory Circuits

Ferroelectric materials are dielectrics in which electric dipoles occur spontaneously and align themselves parallel by mutual interaction. Because of this the dielectric induction versus applied electric field intensity curves for ferroelectric materials show hysteresis loops similar to the curves of flux density versus field intensity for ferromagnetic materials. Thus the designation ferroelectric material signifies the dielectric analogue of a ferromagnetic material rather than a material containing iron.

A number of dielectric materials such as barium titanate, rochelle salt, potassium niobate, and potassium dihydrogen phosphate have been found in recent years to have ferroelectric properties. At the present time barium titanate ($BaTiO_3$) appears to be the most practical of these materials for use in memory circuits. It remains ferroelectric over a temperature range from below 0 degrees centigrade to about 120 degrees centigrade and it will polarize with lower electric field strengths than are required for any other known ferroelectric material except rochelle salt. Rochelle salt has the disadvantage of being ferroelectric only from about −15 degrees centigrade to about +20 degrees centigrade.

Especially prepared single crystals of barium titanate 0.004 to 0.010 inch thick and having an area up to about 0.1 square inch have been used for the experimental memory circuits to be described. These crystals and methods of obtaining characteristics suitable for memory cricuits were developed by J. P. Remeika of the Bell Telephone Laboratories. Silver paste electrodes about 0.025 inch in diameter are placed on opposite sides of the crystals in one or more positions to form individual memory cells. A typical crystal with two sets of electrodes attached is shown in Figure 1.

As will be shown, substantially rectangular hysteresis loops and low coercive forces are desirable in ferroelectric materials to be used in memory circuits. Barium titanate crystals and ceramics which have been prepared for transducer and capacitor application in general are not satisfactory for memory circuits because of their high coercive forces and nonrectangular hysteresis loops.

On Figure 2 are shown 60 and 6,000-cycle-per-second hysteresis loops for one type of barium titanate crystal which is suitable for memory circuits. Hysteresis loops of the type shown are observed by applying voltages proportional to electric field strength and polarization to the horizontal and vertical plates respectively of a cathode-ray tube when a 60-cycle-per-second (or 6,000) electrical field is applied across the ferroelectric material. Circuits for accomplishing this are described by C. B. Sawyer[1] and W. G. Cady.[2] The 60-cycle-per-second loop

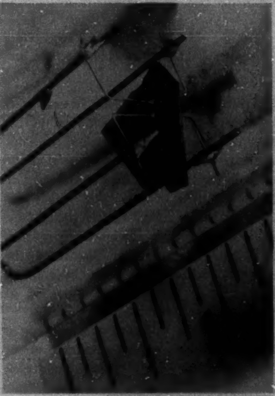


Figure 1. Barium titanate crystal containing two independent memory cells formed by two sets of 0.020-inch diameter plates whose edges are separated by 0.005 inch. The plates or electrodes are formed by firing silver paste spots onto the surface of the crystal. The leads to the plates are 0.003-inch diameter headed wires soldered to the silver surface of the plates

---

Paper 52-309, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted February 20, 1952; made available for printing September 3, 1952.

J. R. Anderson is with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

The author wishes to acknowledge the valuable information and suggestions on ferroelectric materials given by J. P. Remeika, B. Matthias, W. J. Merz, Mrs. E. A. Wood, P. W. Anderson, and J. K. Galt of Dr. S. O. Morgan's Solid State Physics Department.

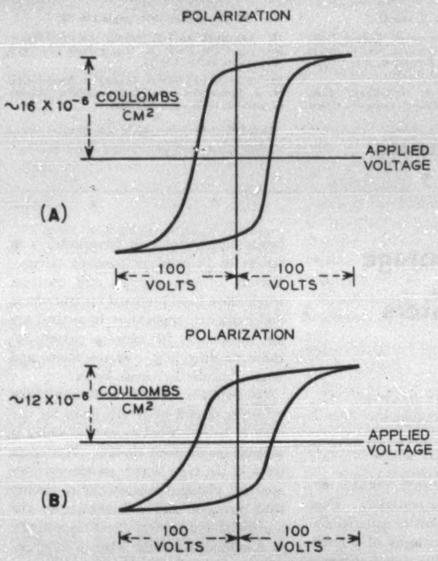


Figure 2 (above). Hysteresis loops traced from an oscilloscope display for a barium titanate crystal 0.0045 inch thick

(A) 60-cycle-per-second loop for 100 volts peak applied voltage
(B) 6,000-cycle-per-second loop for 100 volts peak applied voltage

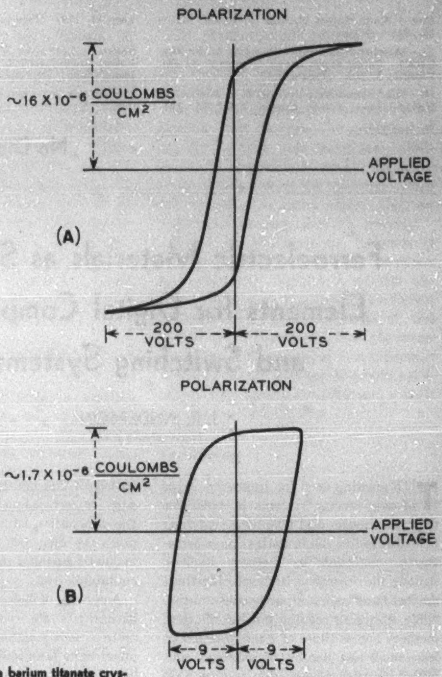


Figure 3 (right). Hysteresis loops traced from an oscilloscope display for a barium titanate crystal 0.0035 inch thick

(A) 60-cycle-per-second loop for 200 volts peak applied voltage
(B) 60-cycle-per-second loop for 9 volts peak applied voltage

illustrated in Figure 2(A) has an extremely high ratio (greater than 50 to 1) between the slope at the sides and the slope at the bottom and top and shows that 100 volts is required to saturate the crystal. Other crystals of this type which saturate with only 25 volts peak applied voltage for 0.005-inch thicknesses are now available. Figure 2(B) illustrates the manner in which the slope of the loop and the remanent polarization decrease at higher frequencies.

It has been found that another type of barium titanate crystal requiring about 200 volts to drive 0.0035-inch-thick crystals to saturation also has a stable low voltage hysteresis loop in the 10-volt range. The voltage at which this loop occurs can be determined by observing the hysteresis loop while the alternating voltage across the crystal is gradually increased from zero. At low voltages the polarization increases very rapidly for small changes in applied voltage. A point is finally reached above which further increase in the applied voltage produces only small changes in the polarization. It is at this point that the loop has been found to be stable enough for operation in memory circuits. Hysteresis loops for a material of this type are illustrated in Figure 3. The hysteresis loop for low voltages is seen to be more rectangular in shape than the hysteresis loop with saturation voltages applied. The ratios of the slopes of the sides to the top of the low-voltage loop is greater than 50 to 1.

While single-frequency hysteresis loops such as those illustrated are useful in predicting the performance of memory circuit materials, they do not give complete information because square pulses containing a wide band of frequencies are applied to the materials in operating circuits.

Ferroelectric materials like other dielectrics exhibit relaxation phenomena which introduce high-frequency losses. The dielectric constant in barium titanate is not affected by relaxation phenomena below about 100 megacycles. However, mechanical resonance phenomena which also introduce losses do occur at much lower frequencies and depend on the crystal dimensions. Relaxation phenomena due to stresses may occur in any frequency range. No adverse effects from these relaxation and resonance phenomena have been observed in the operation of ferroelectric memory circuits. However, it is necessary to have firm connections between crystal electrodes and their lead wires and to allow freedom of motion of the crystals because of electromechanical resonances due to the mechanical resonance phenomena.

The Basic Ferroelectric Memory Circuit

The basic circuit consists of a minute capacitor containing a ferroelectric material as the dielectric in series with an output capacitor which is shunted by a rectifier or resistor as shown in Figure 4. The rectifier, which may be of the germanium diode type, is used only when it is desirable to prevent storage pulses from

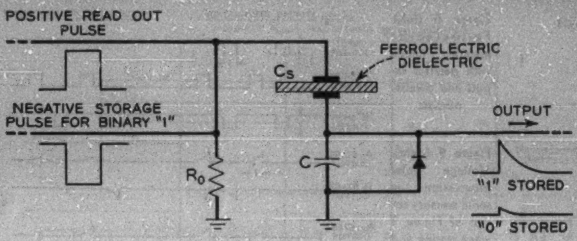


Figure 4. Schematic diagram of basic ferroelectric memory circuit

appearing across the output of the memory circuit. The ferroelectric material is initially polarized in one direction by a voltage pulse (called a read-out pulse) applied across it and the series output capacitor. The application of this pulse clears out any memory of previous conditions in the ferroelectric and restores it to a normal state in preparation for storage. In this state a binary 0 may be said to be stored as no storage pulse is ever applied to store a binary 0. A binary 1 is stored in the ferroelectric by applying to it a voltage pulse whose polarity is opposite to the read-out pulse. The ferroelectric is thus polarized in the opposite sense from its normal state when a binary 1 is stored.

The stored information is read out by applying the read-out pulse across the ferroelectric and output capacitor. The voltage across the output capacitor at the instant the read-out pulse is applied will be high if the ferroelectric presents a high capacitance and low if it presents a low capacitance to the read-out pulse.

As will be shown the ferroelectric does present a low capacitance to a pulse of the polarity of the read-out pulse when in its normal state (binary 0 stored), while in the oppositely polarized state (binary 1 stored) the ferroelectric presents a relatively high capacitance to the read-out pulse. Thus the type of information stored is indicated by the amplitude of the output pulse.

Figure 5 shows the dynamic ferroelectric hysteresis loop involved in the operation of the circuit of Figure 4. The shape of the loop approximates the observed 60 cycles-per-second loop of Figure 2(A). The ordinates on Figure 5 are P, the polarization of the crystal, and the abscissae are V, the applied voltage (equal to the applied electric field strength E times the crystal thickness t). For any segment of the loop, the capacitance of the crystal is the ratio of the change in polarization per unit volume to the change in applied field. If A is the plate area and $V=Et$, then

$$C_s=\frac{A}{t}\frac{dP}{dE}=\frac{AdP}{dV}$$

where C_s is the crystal capacitance. It is observed from the figure that this ratio is large along the steep parts of the segments B to C and A to D, and small over the segments A to C and B to D. In its normal state of polarization (binary 0 stored the ferroelectric material will be in state A. If a positive read out pulse is applied the loop is traversed from A to C and then back to A upon removal of the pulse. The slope of the loop along this part indicates a low capacitance. When a binary 1 is stored in the material by applying a negative pulse, the loop is traversed from A to D and then back to B leaving the material in state B upon removal of the pulse. If a positive read-out pulse is now applied, the material will change from state B to B' during the initial rise of the pulse. B' is the point at which the product of the voltage across the ferroelectric times its capacitance becomes equal to the product of the voltage across the output capacitor times its capacitance. The capacitance at B' is quite high as indicated by the steep slope of the hysteresis loop at this point and most of the voltage of the read-out pulse appears across the fixed output capacitor. As the output capacitor discharges through the back resistance of the diode, the voltage across the ferroelectric increases until point C (saturation) is reached. When the read-out pulse is removed, the loop is traversed from point C back to A and the ferroelectric is again left in its normal state.

The polarities of the storage, read-out, and output pulses can be reversed from those shown in Figure 4 if the direction of the diode also is reversed.

Requirements on Capacitances and Resistance for the Basic Memory Circuit

It is desirable in a memory circuit to have the output voltage when reading out stored information as large as possible compared to the input voltage. It is also necessary to have a large ratio between the output voltage when reading out a stored binary 1 and the output voltage when reading out a stored binary 0. From Figure 5 it can be seen that the output voltage for reading out a stored binary 1 must be of such an amplitude that the loop will be initially traversed at least up to point B' on the steep section. It is thus desirable that the sharp corner in the hysteresis loop be close to the ordinate axis compared to the maximum voltage applied. For the loop shown in Figure 5 this means that the maximum amplitude of the output voltage is limited to about 80 per cent of the read-out voltage. The output pulse can, of course, be increased to a higher percentage of the read-out pulse by applying read-out pulse voltages greater than saturation. For low-voltage loops of the type shown in Figure 3(B) the output voltage is limited to about 30 per cent of the read-out voltage.

The ratio for the output voltage V_1'' when reading out a stored 1 to the read-out voltage V_1 is expressed by

$$\frac{V_0''}{V_1}=\frac{C_s''}{C_s''+C} \tag{1}$$

where C_s'' is the high capacitance state of the ferroelectric and C is the fixed output capacitor.

The ratio of output voltage V_0' when reading out a stored 0 to the read-out voltage is expressed by

$$\frac{V_0'}{V_1}=\frac{C_s'}{C_s'+C} \tag{2}$$

Now let $C_s''=XC_s'$ where C_s' is the low-capacitance state of the ferroelectric. It then follows from equations 1 and 2 that

$$C=C_s''\frac{(R-1)}{(X-R)} \tag{3}$$

where R is the ratio of output voltage

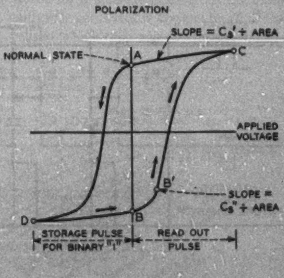

POLARIZATION
SLOPE = C_s' + AREA
NORMAL STATE
A
C
APPLIED VOLTAGE
B'
SLOPE = C_s" + AREA
D
B
STORAGE PULSE FOR BINARY "1"
READ OUT PULSE

Figure 5. Dynamic ferroelectric hysteresis loop

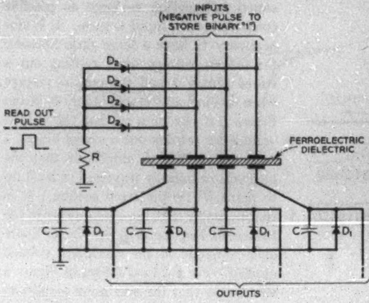


Figure 6 (left). Ferroelectric memory circuit with parallel inputs and parallel outputs

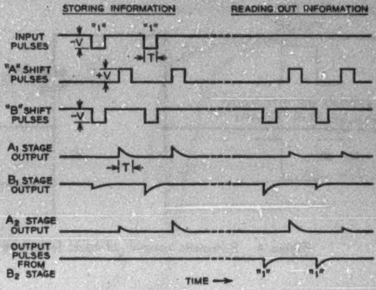


Figure 8 (right). Voltage pulse progression in serial memory circuit of Figure 7 when storing and reading out the binary number 11

when reading out a stored 1 to that when reading out a stored 0.

When the shape of the dynamic hysteresis loop is known, it is possible to calculate the best value for the fixed output capacitor C by using equation 1 and setting V_0''/V_1 to a value of about 0.8 (or 0.3 for low-voltage loops). Then after determining X from the hysteresis loop, equation 3 can be used to determine the ratio R to predict the performance of the circuit. It has been found that when C_s'' and C_s' values obtained from 60-cycle-per-second hysteresis loops are used for calculations with equations 1, 2, and 3, good agreement is obtained with observed results using storage and read-out pulses which are 5 milliseconds or longer. For pulses in the microsecond range, C_s'' and C_s' values must be obtained from hysteresis loops made at much higher frequencies. In the latter case it usually is easier to determine the optimum value of the fixed capacitor C experimentally with an operating circuit.

The value of the back resistance R_B of the diode (or a resistor replacing the diode) shown in Figure 4 is determined by the length of storage and read-out pulses used. It must be low enough in value to allow the essentially complete discharge of capacitor C and charge of C_s during the length of a pulse so that the voltage across the ferroelectric will always rise to the value of the storage and read-out pulses. This value can be expressed by

$$R_B(C_s''+C)\cong T/3 \qquad (4)$$

where T is the pulse length in seconds, C and C_s'' are in farads and R_B is in ohms.

Performance of the Basic Memory Circuit

Basic memory circuits employing material of the type for which the hysteresis loop is illustrated in Figure 2(A) have been constructed which operated satisfactorily on ±30-volt storage and read-out pulses which are 5 milliseconds long. The output voltage when reading out a stored binary 1 is 25 volts and the output for reading out a stored binary 0 is only about 0.6 volt, when using a 0.01-microfarad output capacitor. The value of X (ratio of high capacitance state C_s'' to low capacitance state C_s') realized for this material in operation is about 280.

Circuits operating on the low-voltage loops such as illustrated in Figure 3(B) have been used with storage and read-out pulses 9 volts in amplitude and 5 milliseconds long, and give output pulses of about 3 volts and 0.3 volt respectively for reading out a binary 1 and a binary 0 with a fixed output capacitor of 0.001 microfarad. With 30-volt storage pulses 0.1 microsecond long and 10-volt read-out pulses 3/4 microsecond long, the output pulses are 3 volts after storing a binary 1 and about 1 volt after storing a binary 0. However, the ratio of the areas under these respective pulses is about 10 to 1. For these short storage and read-out pulses the output capacitor required is only about 100 micromicrofarads.

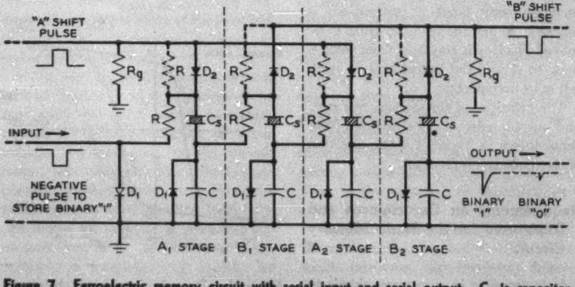


Figure 7. Ferroelectric memory circuit with serial input and serial output. C_s is capacitor with ferroelectric dielectric

As the information stored in the ferroelectric material depends upon internal polarization and not on surface charge, the memory of the material will last for several days even with the electrodes short-circuited during this time. When operating on the low-voltage loop as illustrated in Figure 3(B), it has been found that the output voltage when reading a binary 1 left stored for 16 hours is about one half that observed when reading out a binary 1 stored for only a few minutes. However, when operating with saturation loops as shown in Figure 2, stored binary information lasts several days without change. Tests also have been made with operation on the low-voltage loop to determine if the state of material changes after applying over 10,000 consecutive read-out pulses before storing a binary 1 and then reading it out. There was found to be no appreciable change in the output voltage under this condition compared to alternately storing and reading out binary 1's.

It has been found that several independent sets of electrodes can be placed on the same crystal of barium titanate without appreciable interference between adjacent memory cells. For example, the crystal shown in Figure 1 has two separate memory cells with a separation of only

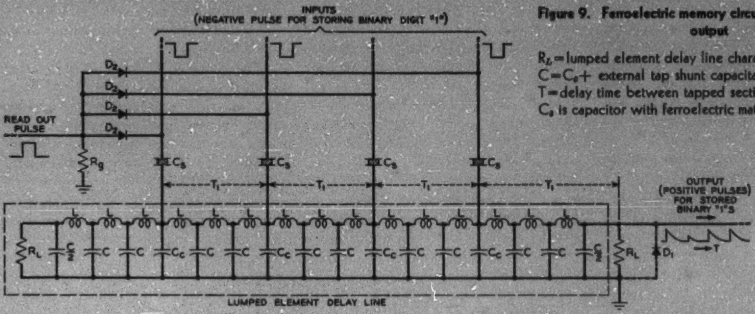

INPUTS
(NEGATIVE PULSE FOR STORING BINARY DIGIT "1")
D₂
READ OUT PULSE
Rg
Cs
T₁
L
RL
C/2
C
Cc
D₁
T
OUTPUT
(POSITIVE PULSES)
FOR STORED
BINARY "1"S
LUMPED ELEMENT DELAY LINE

Figure 9. Ferroelectric memory circuit with delay line output

R_L = lumped element delay line characteristic impedance
$C = C_c +$ external tap shunt capacitance to ground
T = delay time between tapped sections of delay line
C_s is capacitor with ferroelectric material as dielectric

0.005 inch between edges of electrodes on the same surface. Tests on this unit showed that information could be stored and read out continually from either one of the cells without appreciably affecting the information stored in the other cell.

Energy and Power Considerations in the Basic Memory Circuit

The energy required for storing a binary digit 1 in the basic ferroelectric memory circuit (assuming a diode across the output capacitor) is determined by the area of the dynamic hysteresis loop and the electrode area. The energy required for reading out a binary digit 1 is determined by the same factors and also must include the energy required to charge the output capacitor. No energy is required to hold a digit in storage and the energy required for reading out a binary 0 will be very small compared to that required for reading out a binary 1.

When operating with storage and read-out pulses of the order of 5 milliseconds long it has been found that the dynamic or operating hysteresis loop approaches the general shape of the 60-cycle per-second hysteresis loop. On this basis it has been calculated that the energy for storing a binary digit 1 with a 9-volt pulse using a material having a hysteresis loop such as is shown in Figure 3(B) with 0.020-inch diameter electrodes is about 7.4×10^{-8} joule. If binary digits are stored at the rate of 100 per second, the average power consumed will be 7.4 microwatts. The average power required for reading out a stored binary 1 at the same rate is 7.8 microwatts.

The energy required when operating with pulses in the microsecond range can be estimated by assuming that the area of the dynamic hysteresis loop will be proportional to the area of the 60-cycle per-second hysteresis loop in the same ratio as the applied voltage and the effective capacities (C_s'' and C_s') at the side and top of the loop for microsecond operation are proportional to the applied voltages and effective capacities C_s'' and C_s' for 5 millisecond operation. The energy consumption for storing a binary digit 1 with 30-volt pulses will then be about 2.05×10^{-8} joule which would give an average power consumption of 20.5 milliwatts when storing at a rate of 1,000,000 per second. The average power for reading out a binary digit 1 at the same rate with a 100-micromicrofarad output capacitor and 10 volt read-out pulses will be about 2.8 milliwatts.

Circuit Applications of the Basic Ferroelectric Memory

The basic ferroelectric memory circuit of Figure 4 has been extended for use in a number of circuits which might find applications in the digital computer or telephone switching field. In Figures 6 to 11 inclusive are presented five of the most important types. The polarities of storage and read-out pulses as shown in these circuits may be reversed if the polarities of the diodes shown is also reversed.

Parallel Input and Output Storage Circuit

The circuit of Figure 6 is designed to store binary digits arriving on the parallel input leads. Only four storage cells operating in parallel are shown, but any number desired could be added up to the limit of the available space on a given crystal surface. As has been noted, each pair of electrodes on the crystal will act independently of all other electrodes as a separate storage cell. Each of these cells operates in the same manner as the basic memory circuit of Figure 4 and the design requirements on capacitor C and the back resistance of diode D_1 are the same. A negative pulse is applied to each input lead where a binary 1 is to be stored. These pulses need not occur simultaneously. The input information will remain stored until a single pulse from the read-out pulse generator is applied to all of the cells at which time positive pulses will ap-

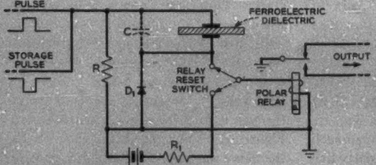
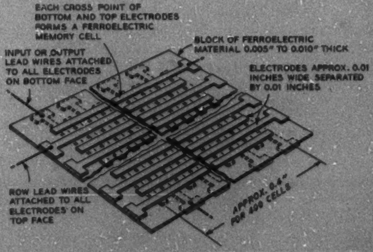

READ OUT PULSE
STORAGE PULSE
C
FERROELECTRIC DIELECTRIC
R
D₁
RELAY RESET SWITCH
OUTPUT
POLAR RELAY
R₁

Figure 10 (left). Ferroelectric memory circuit with relay output

Figure 11 (right). Physical embodiment of a 2-dimensional ferroelectric storage array for 400 binary digits


EACH CROSS POINT OF BOTTOM AND TOP ELECTRODES FORMS A FERROELECTRIC MEMORY CELL
INPUT OR OUTPUT LEAD WIRES ATTACHED TO ALL ELECTRODES ON BOTTOM FACE
BLOCK OF FERROELECTRIC MATERIAL 0.005" TO 0.010" THICK
ELECTRODES APPROX. 0.01 INCHES WIDE SEPARATED BY 0.01 INCHES
ROW LEAD WIRES ATTACHED TO ALL ELECTRODES ON TOP FACE
APPROX. 0.4" FOR 400 CELLS

pear on the parallel outputs of cells where a binary 1 was stored. Application of the read-out pulses also returns the ferroelectric memory cells to their normal condition so that successive read-out pulses will cause only small positive pulses corresponding to binary 0's to appear on the output leads until new information is stored. The diodes D_2 are used to decouple the individual memory cells but still allow them to be driven from a common read-out pulse generator. Their back resistance should be large compared to R to give a large amount of decoupling. Diodes D_1 are used to prevent storage pulses from appearing on the outputs. They may be replaced by a resistor equal to their back resistance if it is not necessary to prevent input pulses during storage from appearing on the outputs. A prototype of this circuit using the crystal of Figure 1 has been operated with 5 millisecond pulses.

Serial Input and Output Storage Circuits

A circuit for storing binary digits in serial form is shown in Figure 7. After the storage period, the digits are pulsed out of the circuit again in serial form. While the circuit, as shown, provides places for storing only two binary digits, additional digits can be stored by adding any number of series stages to the output end of the circuit. One A stage and one B stage are required for each additional digit to be stored. Negative pulses are used to store binary 1's and no input pulses to store 0's.

The first digit to arrive is stored in the A_1 stage, then is transferred to the B_1 stage by the A shift pulse before the next input pulse arrives. The second digit is stored in the A_1 stage at the same time that the first digit is transferred from stage B_1 to stage A_2 by the B shift pulse. Both digits are then transferred by the A shift pulse to the B_1 and B_2 stages respectively. The information then remains stored until B and A shift pulses are again applied at which time the first digit appears at the output and the second digit transfers to stage A_2. The second digit then transfers to stage B_2 and then to the output. The time relations of the various pulses used to store and read out the binary number 11 are shown in Figure 8.

Pulses are prevented from passing backward from the B stages to A stages when B shift pulses are applied by making the resistance of R large compared to the forward resistance of diode D_1. The D_2 diodes are used to prevent coupling between the two A stages and between the two B stages. The back resistance of these diodes should be large compared to the series combination of R and the back resistance of the D_1 diode. The output capacitor C of each stage includes the effective shunt capacitance of the following stages. The value of R_g should be equal to or greater than R to allow at least one-half the output voltage of each stage to be applied to the next stage. Calculations for the values of capacitor C and the resistance $R+R_g$ (which is the limiting back resistance across D_1) are the same as for C and R_0 of the basic memory circuit.

The circuit of Figure 7 can be modified to eliminate one-half of the diodes required by replacing the D_2 diodes with resistances R as shown dotted in on the circuit diagram. The value of R_g must then be low compared to R to prevent coupling between the outputs of the two A stages or between the two B stages. It is also necessary in this modification to double the magnitude of the input and A and B shift pulses, because of the voltage division which occurs when each diode D_2 is replaced by a resistance R.

Serial-type circuits comprising two to four cells have been constructed and operated with 5 millisecond pulses. It has been found that the output voltage of these circuits even when loaded down on the output with a shunt resistance and capacitance equal to that for adding following stages is as high as the output from the A_1 or first input stage. This indicates that any number of stages may be added in tandem.

Delay Line Input Circuit

Figure 9 shows a circuit in which binary digits arriving on parallel leads are stored in ferroelectric memories and then pulsed into tapped points on a delay line. The stored information appears in serial form at the output of the delay line. The parallel memory circuits operate in the same manner as those on Figure 6, capacitor C of Figure 6 being replaced by the effective shunt capacitance C_1 of the delay line at the tap points. A single diode D_1 can be shunted across the output of the delay line when it is desirable to keep input storage pulses from appearing on the output. The voltage pulses on the input and output when a binary number 1011 is stored and read out also are illustrated in Figure 9.

While only four storage cells are shown, more cells can be added in parallel as long as additional sections of delay line are added to provide delay T_1 between each tap point. The total number of delay line sections which can be added is limited by the attenuation of pulses passing down the delay line.

The circuit of Figure 9 has been tested using a Western Electric *D-172597* delay line and single barium titanate crystals tapped in at various points along the line. This delay line has a characteristic impedance of 415 ohms, shunt capacitance C (Figure 9) of 75 micromicrofarads, a cutoff frequency of about 8 megacycles, and a total delay of about 5 microseconds. The crystals used had very high saturation voltages so that 1/2 microsecond pulses 115 volts in amplitude were required to store and read out information. Output voltages of 27 volts and 4.4 volts high respectively were observed when reading out a stored binary 1 and 0 with the crystal connected to a tap point 1.7 microseconds from the output. With a crystal tapped into a point 3.4 microseconds from the end of the line, the output voltages for reading out stored binary 1's and 0's were 23 and 4.0 volts respectively.

Storage Circuit with Relay Output

The circuit of Figure 10 illustrates a ferroelectric storage circuit which operates a polar relay when a binary digit 1 is read out or leaves the relay in an unoperated condition when a binary digit 0 is read out. The output of this circuit consists of open or closed relay contacts which may be used to light a lamp, operate other relays, or perform some other function. Information may be stored in the ferroelectric when the relay is either in its operated or unoperated condition because diode D_1 allows the ferroelectric to be polarized by an input pulse without passing current to the relay. The relay operates when a read-out pulse is applied and a binary 1 has previously been stored in the ferroelectric because the ferroelectric presents a capacitance in this state which is sufficient to cause a large current pulse to flow through the relay When the ferroelectric is polarized in the same direction as the read-out pulse (a binary 0 is stored), application of the read-out pulse will initiate only a small current pulse through the relay due to the very low capacitance of the ferroelectric in this state. Thus, the relay cannot operate when reading out a binary 0.

Because of its "locking" polar characteristics, the relay remains operated after receiving a large current pulse and can be reset only by passing a reverse current through it with the reset switch shown in Figure 10. The capacitance C shown in the figure may be added when the area of electrodes on the ferroelectric is too small to provide sufficient capacitance to operate the relay. The capacitance of capacitor *C* must be smaller than that required to operate the relay so that there will be no operation when reading out a stored binary 0 from the ferroelectric.

A circuit such as shown in Figure 10 has been used to operate a Western Electric type *276D* relay with 45-volt read-out and storage pulses. A shunt capacitor *C* of 0.1 microfarad was used with the ferroelectric because the electrode area of the latter was only about 0.009 square inch. This circuit has always operated successfully even when reading out a binary 1 which had been left stored in the ferroelectric for a period of 1 week.

Two-Dimensional Storage Circuit

The 2-dimensional storage array provides a means for compacting a large number of storage cells into a small area and eliminates the requirement for individual output capacitors and diodes or resistances for each cell. The physical embodiment of this system consists of a large single crystal having a set of parallel-line electrodes on one face perpendicular to a set of parallel-line electrodes attached to the other face as shown in Figure 11. Each crosspoint of the top and bottom electrodes forms a separate memory cell and information independently may be stored in and read out of any of these cells or of any row of cells. A binary digit 1 may be stored in any cell of the array by applying one-half the voltage required to polarize the ferroelectric to one top electrode and an equal but opposite polarity voltage to one bottom electrode. The cell in which the digit is stored will be that formed by the crosspoint of these electrodes. Thus, the full polarizing voltage will appear across that cell but only one-half the polarizing voltage will appear across all other cells in the array. The information is read out of any individual cell by applying a full polarizing voltage (of opposite polarity to that used for storing a binary digit 1) across the cell and an output capacitor shunted by a resistance as in the basic memory circuit. By storing information in all cells of one row simultaneously and reading out information from all cells in one row simultaneously, it is possible to arrange the output circuit so the capacitance of all other memory cells in each column acts as the fixed output capacitor for that column. With this 2-dimensional circuit as many as 2,500 bits might be stored on a 1-inch square crystal of the order of 0.005 inch thick.

This type of circuit requires a ferroelectric material which will not change its original state of polarization whenever positive or negative voltage pulses one-half the amplitude of the normal polarizing pulses are applied across the material. Such a material recently has become available and preliminary tests with this material containing two cells in the same row and in the same column have indicated that it might be possible to extend the 2-dimensional type of operation to a large array as shown in Figure 11.

Conclusions

Even though ferroelectric storage devices are in their early development stages and present access problems in a large array, they show considerable promise as a new method for storing digital information in computers or switching systems. These devices provide a means for storing up to 2,500 bits of information per square inch in a surface only a few thousandths of an inch thick with pulses less than a microsecond long. In addition they offer the possibility of operation from low-voltage (that is, 10 volts or less) circuits such as transistors, consume no power during the storage period, are adaptable to a wide variety of circuit combinations, and provide storage for long periods of time without regeneration.

References

1. Rochelle Salt as a Dielectric, C. B. Sawyer, C. H. Tower. *Physical Review*, American Institute of Physics (New York, N. Y.), volume 35, February 1, 1950, pages 269–73.

2. Piezoelectricity (book), W. G. Cady. McGraw-Hill Book Company, Inc., New York N. Y., 1946, pages 559–60.

No Discussion

Radio Interference Control

CARLETON F. MAYLOTT
ASSOCIATE MEMBER AIEE

RADIO receivers respond to radio transmitters by virtue of antennas which couple them both to free space. As compared with wires, free space is an inefficient coupling medium, and it has erratic efficiency changes. These changes are well known as fading, skip, and sporadic effects, which vary with distance, direction, frequency, and time.

Receivers are subject to many other undesirable influences, including:

1. undesired signals from transmitters using the same, adjacent or image-related frequencies, and those emitting harmonic or spurious frequencies;
2. atmospherics or natural static;
3. celestial radiation or cosmic noise;
4. background noise due to thermal agitation in conductors and shot effect in tubes;
5. a-c hum;
6. microphonic tubes;
7. poor connections;
8. faulty components; and
9. man-made electrical disturbances.

For the purpose of this paper, all but the last of these influences may be disregarded as subject to the control of the radio designer, the radio operator, or nature alone. However, the man-made electrical disturbances are largely subject to the control of electrical engineers, with the aid of others, including radio engineers. Such control may be called "radio interference control."

Certain electrical devices, which use a-c or d-c power, produce radio-frequency byproducts as electrical disturbances in space, even though they are in perfect condition. The Federal Communications Commission in 1949[1] defined such incidental radiation devices as "devices which radiate energy substantially from a point source, which radiations are incidental to the work to be accomplished, and which devices do not require the use of associated receivers." The word

---

Paper 52-300, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted February 25, 1952; made available for printing September 4, 1952.

Carleton F. Maylott is with Scintilla Magneto Division, Bendix Aviation Corporation, Sidney, N. Y.

"radiation" may be considered to include conduction and induction, as all modes of wave propagation are effective at close range, the condition usually associated with radio interference complaints.

For many years such electrical disturbances caused only acoustic noise at audio outputs of radiotelegraph and radiotelephone receivers, and the term "radio noise" was self-explanatory. Upon the addition of new radio services, including facsimile, television, and remote control, the meaning of radio noise was broadened to cover any disturbances causing any undesirable effect at the output of any receiver. With the development of electronic equipment for nonradio functions, the meaning of "receivers" also was broadened.

The term "radio noise" is now largely superseded by the inherently broader term "radio interference." As commonly used in military specifications and elsewhere, radio interference denotes "any electrical disturbance which causes undesirable response or malfunctioning of any electronic equipment." Conducted interference denotes "radio interference that is propagated along circuit conductors," particularly unshielded control and power wires, including both wire-to-wire and wire-to-ground circuits. Radiated interference denotes "radio interference that is propagated in electromagnetic fields, including both induction and radiation components."[1]

These components, in the case of an elementary dipole radiator in free space, are equal at a distance of 0.159 wave length ($\lambda/2\pi$) or $47.7/f$ meters, where f is the frequency in megacycles. The induction component is negligible at more than ten times that distance, and the radiation component is negligible at less than one-tenth of that distance, as measured from the center of an excited dipole. This is because the induction fields (electric and magnetic) diminish more rapidly with distance than the radiation field, which follows an inverse-distance law. The electric field has two components, which follow inverse-square and inverse-cube laws, and the magnetic field follows an inverse-square law; thus the electric field is predominent near the dipole. Near a loop antenna the magnetic field is predominant.

Radio Interference Sources

All electric power sources and loads are potential sources of radio interference, but the common exceptions are: batteries, transformers, heating units, incandescent lamps, and polyphase cage-type motors, when in steady-state use and in good condition. Electrical devices and their circuits have unlimited electrical constants and connection complexities, which determine steady-state values of current, voltage, frequency, power, power factor, and resistance or impedance. These steady-state values have little significance in interference studies except as they relate to rapidly changing values and rapidly repeating changes known as transients and semitransients respectively.

Each make, break, or sudden change in a circuit affects the energy stored in electric and magnetic fields. Most of the energy gained or lost at various points in the circuit either is dissipated in the form of heat or returned at another time or place to the circuit, but some energy is transferred by conduction, induction, or radiation to other circuits and space, if conditions permit. The transferred energy is largely a high-frequency component because coupling conditions do not favor the escape of low-frequency energy. The phenomena of generation and propagation of high-frequency waves under such conditions are fairly well understood, but the practical control of these phenomena is an endless engineering problem.

The problem of radio interference control is greatly simplified by ignoring many radio interference sources which are minor or rare offenders. Thus, it is customary to disregard transients caused by relays, switches, and thermostats, if their duration is a fraction of a second and they recur in not less than 3 minutes. Rotating and sliding contacts, as in slip-rings and third rails, may or may not cause excessive transients and corresponding radio interference, depending upon design and operating conditions.

Semitransient or substantially continuous interference sources, which require expert radio interference control, include the following: arcs, corona, commutators, gaseous conduction tubes, ignition circuit-breaker points, spark gaps or plugs, and vibrators.

Radio Interference Effects

Radio interference passes through a receiver as an impostor or false signal when the true signal is absent and as a parasite when the true signal is present. Thus, the receiver delivers sound and noise, images and blotches, or desired functions and undesired malfunctions. These dual responses reduce the speed of telegraphy, the intelligibility of speech, the appreciation of music, the clearness of images, or the accuracy of mechanisms.

A high interference level or low signal-to-interference ratio lowers the range of a receiver for a given quality of reception from a transmitter of a given power. On the other hand, any reduction of interference level raises the useful sensitivity of the receiver toward an absolute limit which depends only on its design. "The sensitivity of a radio receiver is that characteristic which determines the minimum strength of signal input capable of causing a desired value of signal output."[2]

Such improved receiver performance is equivalent to increased transmitter power and may even permit the use of a less powerful, hence cheaper, lighter, and smaller transmitter.

These advantages are well worth a great deal of radio interference control especially in mobile and portable installations of electric and electronic equipment.

Proper Radio Interference Control

It is desirable, at this point, to clarify some general aspects of radio interference control. The word "control" was chosen to emphasize the idea of limits. The usual word "suppression" has a similar meaning, but is rather vague as to limits, suggesting the thought of elimination rather than reduction.

"Ideal" radio interference control, or elimination, may be defined as that permitting any desired electric equipment to be installed and used in close proximity to any desired electronic equipment without detectable influence of the former on the latter, or vice versa, under any operating conditions. Such control virtually would isolate the potentially interfering electric equipment from the potentially susceptible electronic equipment, regardless of actual spacing; moreover, this perfect decoupling would be accomplished without any adverse changes in characteristics of the electric equipment due to necessary changes in circuit constants.

"Practical" radio interference control falls somewhat short of this approachable but unattainable ideal for several reasons. First, as previously noted, it is practically necessary to ignore many minor or rare sources of interference. Second, there are no ideal suppression components available, such as: 1. capacitors having infinite impedance to all power frequencies and zero impedance to all radio frequencies, 2. choke coils or inductors having just the opposite characteristics, 3. filters containing ideal capacitors and coils, 4. resistors having constant resistance and unity power factor at all frequencies, temperatures, and voltage gradients, and 5. shielding having electrically perfect joints and walls, and causing no electrical effect on shielded circuits. Nevertheless, it is often possible to control radio receiving conditions to such an extent that the signal-to-interference ratio is high and sets no noticeable limit on the intelligibility of signals.

"Proper" radio interference control may be defined as the correct use of suitable suppression components in or on potentially interfering electric devices or their circuits, as conditions permit or require. The choice of components may depend upon cost, purpose, specifications, or technological limits. This means that the most suitable suppression components are not necessarily the best procurable, on a purely radio basis of comparison, when the radio interference requirements are not very severe, but there are other limiting factors, such as competition, dimensions, and electrical effects.

These considerations show that radio interference control at its best is an intelligent compromise between two or more important factors, and that the end result is one of several possible maximum levels of radio interference. Like many evils, radio interference is reducible to a comparatively harmless level with much less effort than that required for its elimination. The word "effort" should not be taken lightly; it takes men, money, materials, and management to determine acceptable radio interference levels for all radio services, to achieve such levels, and to maintain them.

Radio interference control of existing electric equipment is often difficult or impossible, requiring additions or modifications known as field fixes. Better and cheaper results are usually obtainable with new equipment, where the radio interference requirements are considered at the design and manufacturing levels. This puts the burden on fewer people, and avoids tinkering with equipment which is better understood by the manufacturers, who are also better prepared to modify it.

In electric equipment manufacturing, as in other lines, it is necessary to maintain reasonable quality at a reasonable price. If the product is too good, it cannot be sold at a reasonable price or profit; if it is too bad, it must be scrapped. If an electric equipment item is subject to radio interference control, it requires added inspection under quality control procedures, which increases its cost, even though it consistently passes all radio interference tests. If the test limits are hard to meet, the cost is further increased by reason of rework and scrap. In seeking to avoid rejects by redesign, the designer may invite penalties of cost, size, or weight.

Under these circumstances, a radio interference expert has a great opportunity and a great responsibility. He must determine a reasonable interference limit and a reasonable factor of safety. He must assist in the design, redesign, testing, and rework procedures. He must avoid undue hardships on either customer or supplier by permitting or requesting reasonable deviations to existing specifications which prove impractical or unnecessary to meet.

Because it is a military necessity, the Armed Services procure electric equipment which is subject to the best possible radio interference control. Military specifications reflect this need for almost ideal control or negligible interference levels. These specifications are usually difficult but possible to meet. They are often beyond commercial radio requirements and cost limitations, but they have promoted the development of scientific radio interference control, which is needed widely now.

Organized Radio Interference Control

Despite engineering advances in certain circles, radio interference control is relatively limited and unscientific in practice. Progress has been hampered by imperfect knowledge of the problems, poor dissemination of knowledge, pride of personal and group opinions, trade secrets, economic limits, and human inertia. There are surprisingly few experts of surprisingly limited experience. There are serious disagreements among them about ways and means and limits of radio interference control. Test equipment, methods, and specifications are not yet standardized, suitable, and scientific. The situation approaches chaos and can be rectified only by improved organization.

In many cases, several parties are jointly responsible for radio interference control. In the aircraft industry there are many parties: engine and ignition manufacturers, electric and radio equipment manufacturers, and airframe manufacturers and their customers, usually the airlines or the Government. It is difficult to get all of these parties together for co-operative work along radio interference lines, although their experts meet occasionally in groups of two or three.

Several engineering societies and industry associations long have shown interest in radio interference and are well equipped to work with each other and with Government agencies on matters of research investigations, recommended practices, and improved specifications. Some co-operative work has been done, but much more remains to be done.

A critical re-evaluation of existing radio interference limits would be very timely and useful, though difficult. Without realistic limits, many manufacturers and their customers encounter undue costs, delays, and misunderstandings. Undue hardships arise when special ways and means are required to accomplish and test radio interference control, with poor correlation between written and working requirements. Safety factors must provide for depreciation, obsolescence, and tolerances, but they should not be too high, or the corresponding limits should not be too low, just to cover exceptional cases or inexcusable ignorance.

Radio interference control is a major undertaking, requiring good organization, as evidenced by the accomplishments of the Armed Services and the airlines for many years. In these cases, rigid Government specifications and supervision were determining factors. The Government is now ending many serious radio menaces through the intervention of the Federal Communications Commission (FCC). In June 1947 the Commission adopted Part 18 of its "Rules and Regulations Relating to Industrial, Scientific and Medical Services," allocating frequencies and establishing radiated field intensity limits at all other frequencies used by radio services. As of June 30, 1952, all radio-frequency heating equipment must be adapted to constant-frequency operation at an allocated frequency or it must be adequately shielded and provided with power-line filters. By April 30, 1953, miscellaneous equipment will be regulated.

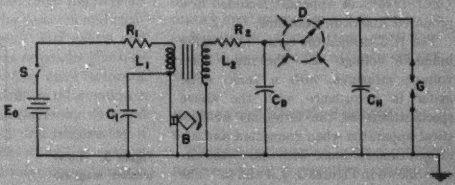


Figure 1. Battery Ignition

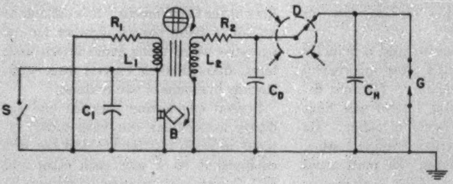


Figure 2. High-tension magneto

The FCC is authorized and required by the Communications Act of 1934, as amended, to "make such rules and regulations and prescribe such restrictions and conditions . . . as may be necessary . . . for the purpose of obtaining maximum effectiveness from the use of radio . . . in connection with safety of life and property, . . . investigate and study all phases of the problem and the best methods of obtaining the cooperation and coordination of these systems." The Commission has refrained from using its full authority to regulate all incidental radiation; as yet it is awaiting action by voluntary agreements between manufacturers, as far as possible.

This means that radio interference control will be enforced, partly by law and partly by public opinion. In areas where the latter fails, the former may take over. Both methods of enforcement have obvious disadvantages. The possibility of legal control poses both an opportunity and a threat to industry. Fortunately, some manufacturing groups have already recognized that fact.

Organized radio interference control or co-ordination began in the early years of radio broadcasting. The electrical utility companies soon recognized the effect of radio noise on customer good will and radio receiver load. In 1924 the National Electric Light Association appointed a committee to study radio noise problems. In 1930 the National Electrical Manufacturers' Association (NEMA) took similar action. In 1931 the NEMA joined with the Radio Manufacturers' Association (RMA) and the Edison Electric Institute to form a Joint Co-ordination Committee (JCC) on Radio Reception. The JCC report of 1940[3] was the basis of the American Standards Association American War Standard *C63.1–1946*[4] and Appendix A added to a proposed American Standard, *C63.2*.[5]

The JCC defined the term "radio noise influence voltage" as "the radio noise voltage measured with a radio noise meter in accordance with the above specifications at the terminals of electrical apparatus when connected and arranged in accordance with one of the circuits shown in Figures 2, 3, 4 and 5." The Committee covered the measurement of line-to-line and line-to-ground noise influence voltages, which, together with power source and load impedances, govern the values of radio noise influence currents. The effect of these voltages upon a radio receiver was truly said to vary with the coupling of power line and antenna, and with the susceptiveness of the receiver. The meaning of "susceptiveness," as used in the JCC report, was not defined clearly but the term "susceptibility," which now is commonly used, has a distinct meaning which will be discussed later in this paper. Values of radio noise influence voltage, as defined here, do not necessarily agree with conducted interference voltages measured according to various Government specifications.

In 1946, after several years of preliminary studies, a joint committee of the RMA and the Society of Automotive Engineers (SAE) established tentative tolerable radio interference limits, suggested means of meeting these limits, and recommended that each vehicle manufacturer should meet the requirements by January 1, 1948. At the same time, the Board of Directors of the Automobile Manufacturers' Association (AMA) recommended to all vehicle manufacturers that they meet the SAERMA standard by altering ignition installations, if necessary, by the proposed date.

Susceptibility and Standards

As commonly understood, the susceptibility of equipment is the ease with which interference causes it to produce undesirable response or malfunctioning, and usually is tested by subjecting an external circuit, except an antenna circuit, to a specified radio-frequency voltage, modulated and unmodulated, throughout a specified frequency range. The susceptibility of a receiver may vary with its bandwidth, frequency, function, linearity, modulation, overload threshold, and sensitivity; however, the chief factor is the coupling between internal and external circuits, as affected by component placements, wiring arrangements, capacitor by-passing, filters, and shielding.

As conceived by the author, the "potency" of a potential interference source may be defined as its inherent ability to disturb a receiver (or other electronic equipment) of arbitrary susceptibility, and it depends upon wave factors of current and voltage at the source, including amplitude, duration, wave form, repetition rate, and regularity of recurrence. The amplitude, duration, and wave form of current and voltage change during propagation from source to receiver, thence through antenna circuits, frequency selection amplifiers, perhaps limiters, frequency converters, detectors or discriminators, untuned amplifiers, and reproducers.

The author suggests the use of the term "potency" to denote differences between electric equipment corresponding to differences in susceptibility of electronic equipment. Susceptibility might be measured by the amount of disturbance at the output of electronic equipment when it is exposed to an arbitrary weak but normally useful signal of arbitrary intensity. Potency might be measured by the amount of disturbance at the output of an arbitrary receiver (of standard susceptibility) when it is exposed to the full amount of interference delivered by the source in question.

Unfortunately, susceptibility measurements are not highly developed like conventional receiver measurements, and potency measurements are also inadequate, reflecting partly the variable susceptibilities of noisemeters or receivers. Potency can be measured to some extent by the use of oscillographs and transient analysis, but these are largely research tools of limited utility.

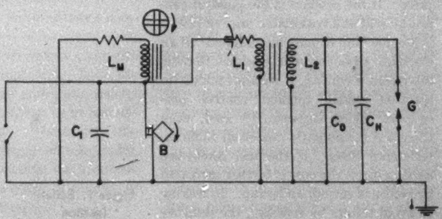


Figure 3. Low-tension magneto

Radio-nuisance values, aural annoyance values, or interference damages, as observed, are complex functions of the characteristics of the interference sources, signals, receivers, and observers. As usually measured with noisemeters, they also depend upon the calibrating means, usually a random noise source or a standard signal generator. Special noise generators having pulse outputs or approximately trapezoidal wave forms more closely simulate common interference sources.

Noisemeters measure conducted and radiated interference in terms of average, quasi-peak, or peak voltages, or at least two of these three responses. They do not measure rms voltage, which is the true measure of power in a resistance load, regardless of frequency and wave form, and is the preferred function in electrical engineering. Quasi-peak readings are preferred in radio interference specifications. They give weight to both peak values and repetition frequencies, but are affected by wave form. They depend upon charge and discharge time constants in meter circuits, as adopted by several engineering groups in America and Europe. The time constants vary widely from one noisemeter to another due to differences in specifications and the degree of conformity achieved by manufacturers.

Various noisemeters, receivers, signal generators, antennas, probes, and accessories have been acquired by many radio interference laboratories, but there is still need for recognized standards of interference potency and susceptibility, in other words, standard interference sources and receivers, with calibrating equipment. The latter would provide a common basis of measurement whereby interference could be evaluated and compared before and after making improvement in offending and affected circuits. Radio interference control then would be more scientific.

There are many serious objections to the use of noisemeters rather than receivers for field and shop testing of radio interference levels. Noisemeters are more complicated, costly, and delicate. Using dry batteries, they require frequent, slow, and expensive servicing. Using auxiliary power supplies, they do not operate with the same amount of conducted interference at the power entry, which is often coupled to the antenna-to-ground circuit. Reading quasi-peak carrier levels, they do not indicate relative effects at the audio output circuit. Noisemeters are not available in quantity in time of local or national emergency. As arbitrary receivers, noisemeters are not designed with typical receiver characteristics, for example, in regard to bandwidths and sensitivities. Early noisemeters were too insensitive; some of the latest exceed receiver capabilities. Some noisemeters are heavy and almost unportable; some are too large for use in tight quarters. None are provided with switches for transferring pickups to the far ends of several long shielded leads, as needed in engine test cells and aircraft installations.

Figure 4. Radio-shielded room, 8 feet high, 10 by 20-foot floor

On the other hand, available inexpensive audio output meters giving average or rms response, or easily built special meters giving quasi-peak response, will permit the calibration of available receivers as radio-frequency noise meters. Specifications based upon this principle have been successful. Even an uncalibrated receiver with any convenient output indicator can be very useful in demonstrating and eliminating interference effects under actual working conditions, as installation engineers have discovered. The same output indicator may be applied to every receiver in an installation by means of jacks or a switch.

Despite the criticisms of noisemeters for routine testing in the field or shop, the author feels that they should have a place in every well-equipped radio interference laboratory as tools for research and engineering test purposes. The values and ratios of their various responses give useful information regarding wave factors of waves as received, but this information is not as complete as oscillographic data and would be more useful and trustworthy if it included rms values as well as average and peak. The traditional emphasis on quasi-peak response is rooted in correlations obtained with an average listener response to average broadcast program material under average interference conditions prevailing on broadcast receivers some years ago. Since quasi-peak time constants have varied from country to country and from time to time, and because present conditions of reception are different and more varied, it is doubted that any set of time constants will prove entirely satisfactory on a permanent international basis as applied to all kinds of radio service and interference sources. The author, therefore, favors readings of average, rms, or peak times repetition frequency. Any given wave can be analyzed to determine values and ratios of average, rms, and peak, although this derived information does not exactly define the wave. But, quasi-peak values give almost no analytical information, and are not derived readily from given wave forms.

Interference waves have so many variable wave factors that several different interference sources may give identical readings, regardless of the type of response selected, without being identical in radio nuisance values. Peak voltages, unrelated to repetition rate, may be identical without causing equal nuisance to music. Triggering devices, like remote-control mechanisms, may respond to occasional peak voltages, even when average, rms, or quasi-peak voltages are negligible.

Engines and Electric Systems

Electric-spark-ignited internal combustion engines serve countless commercial,

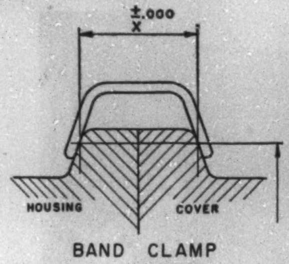


Figure 5. Band-clamp joint

military, and pleasure purposes in aircraft, automotive, marine, and stationary installations. Such engines require either battery or magneto ignition systems. Engine-driven magnetos supply electric power to their own ignition circuits and occasionally are modified to supply additional lamp loads. The independence of the magneto and its superior high-speed performance make it the universal ignition source in aircraft and other critical applications. The magneto also serves most engines not having electric self-starters or heavy electric loads.

There is often a need for stand-by energy to operate a starting motor, lights, and other electric accessories while the engine is at rest. This need calls for a storage battery, hence a d-c generator, a cutout relay, a voltage regulator, often a current regulator, an ammeter, fuses or circuit breakers, and switches. The availability of this complete electric system permits the use of a battery ignition system, having somewhat lower cost, in general, than a magneto ignition system.

The existence of a d-c electric system does not necessarily preclude the use of a magneto ignition system, an a-c electric system, or both, especially in large installations where their advantages are not outweighed by economic considerations. Many mobile or portable electric systems contain inverters, a-c generators, or both, which supply 400-cycle a-c power to loads in the order of 100 watts to 10 kw.

The importance of engines and electric systems is obvious and their dependability is excellent but the importance of radio installations is not so obvious and their dependability largely depends upon their electrical environment. The improvement of this environment is the object of radio interference control.

Ignition Interference

Uncontrolled ignition systems usually cause radio interference when they are located near radio receivers, as in airplanes, motorboats, and vehicles. Ignition interference is seldom detectable at distances beyond 1,000 feet or 1/4 mile, as it usually arises in local, low-power, semishielded circuits. Most spark-ignited engines are used in mobile service, and seldom are exposed for long operating intervals in close proximity to fixed radio stations. However, at certain times and places, traffic density and location is such that serious interference occurs. On mobile units equipped with 2-way short-wave radio, ignition interference must be controlled. In other cases, more or less control is advisable, depending upon circumstances, and may cost anywhere from a few dollars to the price of the engine.

It is interesting to recall that radio communication was largely dependent upon sparks for many years until the advent of dependable high-power vacuum tubes, and that, more recently, sparks assisted in the development of radar. Though sparks have had occasional employment in radio work, they always have been indispensible in internal combustion engines except diesels, which have fewer applications.

An engine is a vital part of any transportation unit and must have a dependable ignition system; its performance cannot be sacrificed to radio interference control. This does not mean that control is impossible, but it means that careful compromises are necessary and that something less than ideal control may be obtainable. Also, because the location of the engine is inflexible, and perhaps unfortunate from a radio viewpoint, somewhat greater interference may arise from the ignition system than from other sources of equal or greater power.

There are two general types of ignition interference control, which may be called the vehicle type and the aircraft type or the suppressor type and the shielding type respectively, although there are exceptions and hybrid cases. The first type also may be called the SAE-RMA-AMA type, although it originated long before the work of the vehicle radio interference committees. The second type may also be called the military type although sometimes it is used in commercial applications (especially in aircraft), and sometimes it is not used in military applications. Both types are as old as mobile radio.

According to SAE recommended practice, radio-noise suppressors are special resistors having nominal values of 5,000, 10,000 or 15,000 ohms, as measured under dynamic conditions.[6] They are connected in series with high-tension spark circuits at either or both ends as required, normally one in the distributor center-lead or tower and one at each spark plug. When resistor-type plugs are used, external suppressors are unnecessary.

The spark gaps between distributor electrodes and spark-plug electrodes are in series and conduct current from the secondary of the ignition coil when the voltage of the latter is sufficiently high. The peak voltage at which each gap breaks down is largely dependent upon the product of gap distance and air density, but the peak current depends upon the applied voltage and series resistance, if gap resistances are disregarded as small. Suppressors limit the current to the order of 1 ampere, as their resistances in ohms are of the same numerical order as the applied voltages in volts. However, in the absence of suppressors, the peak current may be in the order of 100 amperes, as the spark plug leads behave as charged capacitors and the resistances of the gaps are major parts of the circuit resistance. Due to the high peak current, the persistence of current, and the limited energy dissipation in the circuit, the unsuppressed ignition system causes more erosion at the spark-plug electrodes and more radio interference at certain frequencies. But it should not be inferred that suppressors are cure-alls. In practice their benefits are often very small, and they tend to drop the voltage available to damp or fouled spark plugs and raise the voltage at places where it is undesirable or unnecessary.

In past years suppressors having d-c resistances from 15,000 to 30,000 ohms were tried. They were abandoned in favor of lower values for various reasons including: 1. poor ignition performance, 2. not much better suppression, at best, and 3. poorer suppression, as a rule. The design of a good suppressor is difficult. It must have exceedingly low distributed capacitance, of the order of 1 micromicrofarad maximum (1,590 ohms reactance at 100 megacycles). It must be stable electrically to high voltage gradients, stable mechanically to shock and vibration, and stable

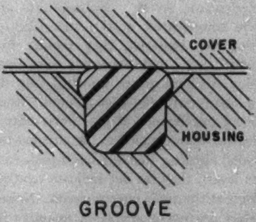


Figure 6. Groove and sealing strip (wire and cotton braid on rubber)

thermally to temperature changes and extremes.

Suppressors do not suppress peak voltages on high-tension leads; they do not affect the distributed constants of these leads nor correctly terminate them to avoid standing waves. The leads are still able to radiate efficiently, especially at certain frequencies for which the lead lengths are resonant. It is thus wise to keep the lead lengths as short as possible, close to grounded metal, within metal brackets or channels where convenient, and far from rods, tubes, and wires which may lie in the induction field and thus enhance radiation.

The low-tension side of an ignition system should not be neglected, as it contains contact-breaker points which spark more or less, despite primary capacitors. Also, the high-tension side is coupled to the low-tension side through mutual inductance and stray capacitances in the ignition coil and the contact breaker-distributor assembly. A suitable by-pass capacitor, preferably of the coaxial or feed-through type, should be applied to the battery terminal of a battery ignition coil. For greater effectiveness over a wider frequency range, a filter with shielded input lead may be required. In either case, radio-frequency currents are largely confined to the ignition components, and no further treatment of the battery wiring should be required, if there is little pickup from the high-tension wiring and if there are no other interference sources.

In most engines there are two other interference sources: the generator and the voltage regulator, one containing a commutator and the other containing a vibrator. Capacitors are applied to the generator armature terminal and to the regulator battery terminal. Better yet, the generator armature and field wires can be shielded jointly or separately as they run to the regulator, which also should be shielded and by-passed or filtered.

In automotive units there are also a number of motors, as in fans, heaters, sirens, and windshield wipers, which should be by-passed. Capacitors often are applied at other points like the ignition switch (coil side), the ammeter, the dome-light, the stop-light, and so forth, which are not interference sources, as this procedure avoids long dead-end leads and loops which help to radiate interference picked up elsewhere. The effect of the latter capacitors is virtually to ground many points at radio frequency, thus altering or raising the natural resonant frequencies of the wiring to frequencies other than those of adjacent radio receivers.

This frequency-shifting procedure may reduce interference at certain frequencies only to increase interference at other frequencies, and is not ideal for wide-band or very-high-frequency control.

If suitable capacitors or filters and metallic shielding are applied to interference sources and circuits, the problems of radio interference control are much simplified and the results are most satisfactory. The location of equipment, the routing of wiring, and the bonding and grounding of metal parts become unimportant, from an interference standpoint.

Shielding is the most effective means of controlling ignition interference, as demonstrated by more than 25 years of aircraft radio experience, but shielding is still largely avoided in automotive practice as too expensive. There is an unfortunate tendency, on the part of many engineers, to charge shielding expense entirely to the ignition system or engine, when it is truly chargeable to radio economies and effectiveness; however, it must be admitted that shielding provides many incidental benefits such as reduced fire, lightning, shock, and explosion hazards, and mechanical protection against dust, moisture, solvents, and vibration.

Regardless of the type and location of engine, ignition interference now can be controlled within acceptable limits by proper design, manufacture, installation, and maintenance of the ignition system. Nevertheless, ignition interference is presently a serious problem connected with private aircraft, motorboats, motor vehicles, stationary engines, and tractors. This interference occurs especially on the newer, higher-frequency radio services. This can be explained on the basis of ignition lead lengths. For an air-insulated wire, the lowest critical frequency, representing one-quarter wave length resonance, can be determined from the approximate relation (within 2 per cent)

$$f = 3000/L, \text{ megacycles}$$

where L is the lead length in inches.

Thus, for leads extending from 12 to 120 inches in length, the critical frequencies run from 250 to 25 megacycles. For one-half wave length resonance, this constant is doubled (6,000), and the critical frequencies run from 500 to 50 megacycles. For other modes of oscillation, the critical frequencies are still higher.

Since ignition pulses have an infinite frequency spectrum and the circuit oscillates freely, with poor damping, it follows that there are many critical frequencies and no dead spots. Even in the best shielded and filtered systems, leakages at two or more of these frequency ranges may be detectable. The predetermination of critical frequencies is not as simple as the relation just given indicates. There are many lead lengths in a single ignition system, as a rule. Many of the leads are not air-insulated in a strict sense, especially when shielded. The velocity of wave propagation is less than the velocity of light, depending upon surrounding medium, and the phase-shift constant is not directly proportional to frequency. In a shielded system, there are two sets of critical frequencies, one corresponding to the inner circuit and one corresponding to the outer surfaces.

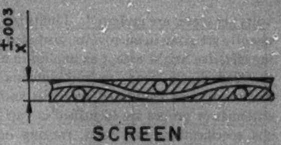


Figure 7. Aluminum screen gasket (rubber-impregnated, 24-mesh)

Government Radio Interference Specifications

Much can be learned from Government radio interference specifications, of which there are many, quite different in details but generally covering these points: use of suppression components covered and not covered by subsidiary specifications, optional and mandatory design features, qualification and production test conditions, test equipment, methods, and limits.

The limits are usually of two kinds: 1. conducted interference voltage between wires, or between a wire and ground, and 2. radiated interference field intensity at a given distance, or probe-test voltages at practically zero distance. Both limits vary according to a curve or table over various frequency ranges. The conducted interference limits are usually harder to meet. They apply to filtered or by-passed devices using common power supplies, and are not applicable to magneto ground leads, which normally are shielded but frequently are unshielded and filtered. The radiated interference limits are more significant as they apply to all devices, and closely simulate severe working conditions.

Radiated interference field intensity is measured in microvolts (quasi-peak) per meter effective height (or length) of an antenna located near an interference source. Distances of 1, 3, and 5 feet are preferred; limits from 2.5 to 15 microvolts per meter are preferred. Limits frequently are given in microvolts, disregarding effective height except as implied in a specified dipole, loop, or rod antenna. Noisemeters or calibrated receivers and antennas or probes are identified in various specifications and tests, ranging in frequency from 14 kc to more than 1,000 megacycles.

Each of the Armed Services has two or more radio interference specifications to cover various equipments and installations. These are amended and superseded often, but earlier issues may apply to new contracts for materials or products for replacement or repair, and to current contracts, unless amended to authorize such changes. Deviations occasionally are requested and granted when the limits prove impractical and unnecessary to meet, or when there appears to be insufficient spread between average test results and limits.

Radio interference limits and tests appear in various places: in detailed paragraphs of specific equipment specifications, in detailed paragraphs of general equipment specifications to which the former may refer, and in subsidiary specifications which deal only with radio interference and may or may not apply to several items of equipment and/or kinds of installation. This maze of specifications and the resulting chaotic conditions are attributable to many causes, including: 1. differing attitudes of the various Services and their subdivisions, 2. differing needs, 3. changing needs due to radio progress, 4. changes in test equipment and methods, 5. improved designs of electric and electronic equipment, and, last but not least, 6. progress in the art of radio interference control. Fortunately, several groups of radio interference experts are now at work in Government and industry circles on difficult and important problems of standardization.

Competitive bidders are often unfamiliar with Government radio interference specifications. They may bid too low, expecting to use standard or special equipment which meets other specifications. Then they acquire a salvage problem and the worse problem of obtaining suitable equipment at greater cost and delay. Sometimes the unsatisfactory equipment can be reworked suitably at the wrong time and place; sometimes no satisfactory equipment can be obtained and deviations must be requested after the contract is signed. Losses and late payments, which are sometimes worse, can be avoided if expert advice is sought in advance of committments, but radio interference experts are scarce, even in very large organizations, and bid requests give few, if any, hints of trouble.

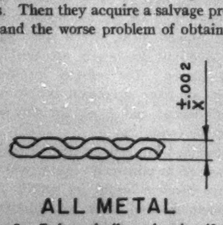


Figure 8. Embossed all-metal gasket (0.012 aluminum, 0.022 minimum by 0.045 wave)

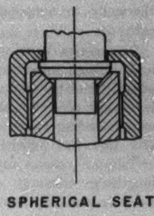
Custom-made equipment is costly because of its special design, limited market, and possibly high rejection ratio. Rejects may be reworked at times, but may not be saleable to another customer because they are misfits. Even if they fit, they may only sell at a loss, as they are better than the average customer demands.

Suppliers need special test equipment, even when the Government loans equipment. Assuming that only two test stations are provided, one in an engineering location and one in a production area, and that both rely on the same spare equipment for emergency, it follows that three sets of equipment meet minimum needs. Suitable calibrating means must be available also at all testing times and places.

As industrial areas are seldom interference-free, several radio-shielded test rooms may be required. Each room must have light, heat, well-filtered power, ventilation or air conditioning, and possibly compressed-air lines. The compressed air is useful for cleaning operations, simulating engine compression on spark plugs, and for driving radio-quiet air motors on test fixtures. Such facilities, see Figure 4, can cost easily from $10,000 to $50,000, but are not amortized easily on small equipment orders.

The maintenance of radio test facilities is also expensive. Due allowance must be made for breakdowns, depreciation, expansion, obsolescence, and frequent servicing. The full-time services from two to ten people, including several radio experts, are required normally. Skilled observers need from 2 hours to 2 days for thorough tests, according to the complexity of the specification and equipment used and tested. These time requirements perhaps are justified for qualification tests and final inspection of large assemblies such as aircraft, but they are uneconomical for 100-per-cent inspection of small items, such as magnetos and motors. It is unlikely that specifications will be simplified very much within the foreseeable future, as the present trend is to the contrary, due largely to increasing frequency range. Thus the small-equipment manufacturer is forced to develop simple, special tests and test-fixtures, maintaining proper correlation with the basic tests.

Figure 9. Spherical (ball-cone) seat, conduit joint



The development of special production tests takes much engineering time and talent, as the tests must be quick and simple, yet reliable, but this results in the saving of much inspection time and talent. Time also can be saved by the simultaneous use of several sets of equipment pretuned to different critical frequencies, but this increases the equipment cost. Several workers are often desirable, as they divide the time, labor, and skill of setting up, testing, and dismantling equipment.

Source inspection, on the basis of actual or simulated installations, is usually desirable and possible. The standards of an equipment supplier must be defended occasionally at the customer's test stand, hence the supplier's test results must be on the safe side of the customer's test results but well correlated with them. This brings up several difficulties. On final installations, such as vehicles, there are several potential sources of radio interference, of which one equipment supplier may be poorly informed, and over which he has no control. If there are faulty designs or applications of any equipment in the final installation, an innocent equipment supplier may have trouble in clearing his responsibility for excessive radio interference. For example, the author often has observed the ruination of well-shielded aircraft ignition systems by the addition of unsuitable switches, ground conduits, starting coils, and filters.

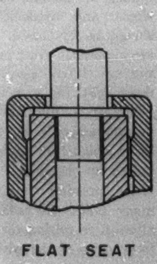


Figure 10. Flat seat, conduit joint

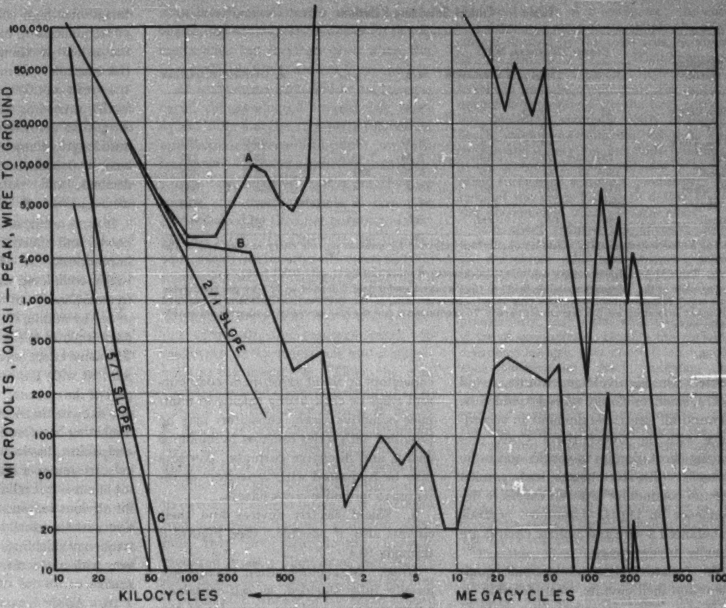


Figure 11. Curves, battery ignition, conducted interference voltage

A: Uncontrolled circuit, 6-volt battery

B: 0.25 microfarad feed-through capacitor added

C: Best available filter added; system shielded, as in A and B

Radio interference limits often govern cost, space, and weight limits, which largely govern competition on equipment of comparable quality. Designers dislike large safety factors which only keep their products safely out of competition. Such being the case, it is necessary at times to work rather close to specified limits, placing great reliance on test instruments and calibrating facilities. Under these conditions, discrepancies in readings of calibrated receivers and noisemeters cause near panic. Signal generators and receivers form a team of reasonable accuracy for radio interference control, being dependable well within a 2-to-1 ratio or 6 decibels, as observed by the author. However, noisemeters vary at least 2 to 1, and often 10 to 1, as a function of time, multiplier-dial setting, make, model, and serial number. These vagaries usually defy the best efforts of laboratory personnel and of factory service departments.

Unfortunately, there are other variables: human errors, differences in test setups and pickup locations, and changes in area noise and background. Specifications are seldom clear and complete on these matters, and leave doubts regarding borderline cases and tolerances. It is important to know when instruments need calibration, or other conditions are abnormal. If false limits are trusted, trouble is sure to follow: some good material is rejected as unfit for shipment, or some bad material is accepted, only to be returned after shipment.

Experience has shown that there are no simple foolproof methods of designing electric, electronic, and engine equipment to meet any or all radio interference specifications, but some manufacturers have built up extensive backlogs of empirical information, and the Government agencies are assisting wherever possible. Radio interference control is not yet very scientific, but good progress has been made. Until better and more universal control has been achieved, it often will be necessary for the Government to modify, or permit temporary deviations to, present radio interference specifications.

Recommended Practices

The following practices are recommended as means of achieving and maintaining radio interference control:

1. Replace interfering devices by noninterfering devices, where possible.

Hints: Use polyphase cage-type motors instead of a-c or d-c commutator-type motors. Use incandescent lamps without flashers or caterpillar-sign mechanisms instead of uncontrolled arcs, fluorescent lamps, or neon lamps. Use any kind of heating instead of high-frequency heating, unless radiation from the latter is well below 10 microvolts per meter at 1,000 feet.

2. Modify interference sources to avoid frequent prolonged transients.

Hints: Clean contacts in relays and thermostats. Provide adequate hold-in force from gravity, magnets, or springs, removing excessive friction at pivots. Avoid chattering motor-starters and control devices.

3. Eliminate static discharges due to friction.

Hints: Ground all machinery. Apply metal grounding brushes or chains to moving material or parts from which discharges occur, for example, belts, paper, textiles, and rubber-tired vehicles. Maintain relative humidity at 40 per cent or more (which nature seldom provides in winter). Avoid moving very dry materials.

4. Eliminate leakage currents and corona.

Hints: Keep all insulation clean and dry. Maintain insulation resistance of at least 1 megohm on low-voltage devices and at least 1 megohm per kilovolt on high-voltage devices wherever practi-

Table I. Critical Shielding Relations

Shielding Quality*	Copper Thickness, Mils† 15.9 Kc	159 Kc	1 Megacycle	z/d	j_o/j_s	A, Decibels	B, Degrees
Skin depth	20.7	6.55	2.61	1.00	2.72	8.686	57.3
Poor	31.6	10	4	1.53	4.62	13.3	88
Thin-Thick	47	15	6	2.30	10.0	20.0	132
Fair	79	25	10	3.84	46.8	33.4	220
Good	130	41	16	6.28	531	54.5	360
Excellent	316	100	40	15.3	4.7×10^6	133	880
Ultimate	5800	1850	740	28.4	2.23×10^{12}‡	247	1630
Ideal	Inf.	Inf.	Inf.	Inf.	Inf.	Inf.	Inf.

* The author's definitions of shielding quality are indicated by underlining one entry in each row of the table.

† See Table II for relative copper-equivalent thicknesses of other metals (1 mil = 0.001 inch).

‡ The ratio of the "strongest possible field" to the "weakest useful field," defined as 22.8 kv per centimeter rms and 1 microvolt per meter rms, respectively, corresponding to the uniform-field dielectric strength of standard laboratory air (25 degrees centigrade, 76 centimeters) and the field intensity of a barely detectable radio wave.

cable. On high-voltage circuits, avoid sharp, small, or closely spaced conductors. Ground all electric equipment in accordance with the Underwriters' Code, preferring short circuits to shocks and radio hazards. Provide bonds and parallel ground conductors to avoid excessive dependence on joints. Grounded neutrals in balanced 2-wire and 3-phase circuits are usually advantageous.

5. Isolate electric and electronic equipment and their circuits, where feasible.

Hints: This procedure is a natural decoupling method, where space permits. There is no reason to invite radio interference deliberately by needlessly close spacing of potent and susceptible equipment.

6. Orient equipment and wiring so that surrounding fields have minimum effect.

Hints: There are three possible axes of rotation in 3-dimensional space. Assuming that the axis of one device is fixed, another device can be rotated 90 degrees from that axis in either of two remaining ways. Like the previous case, this is a natural decoupling method.

7. Minimize external fields by locating related parts and wires close to ground and close to each other.

Hints: Lower, closer parts use less wire in smaller loops having less self-inductance and more or less self-capacitance (length/spacing ratio variable). Mutual inductance and capacitance to other wires normally decrease.

8. Modify electric circuits and devices to subdue transients, suppressing peak currents and/or voltages, and increasing their time duration.

Hints: Steady-state conditions are sought, therefore, anything that lowers inductive voltages ($e=Ldi/dt$) and capacitive currents ($i=Cde/dt$) is advantageous. Coils may be provided with shunt capacitors or shunt resistors or eddy-current rings or close-fitting shields to minimize inductive kicks; capacitors may be provided with series resistors to minimize charge and discharge currents; contacts may be by-passed with capacitors or resistors to minimize break effects.

9. Shield electric devices and their circuits also, if possible. (See Figures 5 through 10.)

Hints: Magneto ignition circuits are usually fully shielded, as self-contained power systems, and are examples of the best radio interference control despite their high potency.

Common thicknesses of common metals as found in sheets, tubes, and castings, provide practical shields having 40 to 120 decibels attenuation at the lowest common radio frequencies. Tables I and II and the Appendix will assist in the selection of suitable metals and wall thicknesses. There are several precautions which must be taken:

1. Avoid flexible metal hose of little effective wall thickness, large porosity, or loose construction (as in *BX*).
2. Avoid nonmagnetic stainless steel or other alloys having low conductivity—permeability product.
3. Avoid electrically imperfect joints by using good practices and conductive-gasket materials such as illustrated in Figures 5–10.
4. Avoid openings of any size, shape, and situation in radio shielding assemblies.

If openings are required for drainage, pressurizing, or ventilation, they should be few in number and as small as possible. If they exceed 1/8 inch in diameter or maximum dimension, they should be covered with fine-mesh metal screens or perforated sheets which make excellent electric contact with housings at their perimeters. In lieu of holes, metal tubes of minimum diameter and of at least two diameters length may be used. Such tubes offer high constant attenuation at all frequencies up to critical frequencies in the order of thousands of megacycles (for example, 1/2 inch = 7000 megacycles; 1/4 inch = 14,000 megacycles, and so forth), according to Linder's general law for piston attenuators (for example, per unit length/diameter ratio coplanar coils have 32.0 decibels, coaxial coils have 66.6 decibels, and mutual-capacitance electrodes have 41.8 decibels).

If slots occur, as in the sides of hinged hoods, and effective wiping contacts are impractical, the shielding may be virtually completed at high and very high frequencies (3 to 300 megacycles and more) by adding close-fitting baffles or capacitor plates to the hood or housing sides. The same effect occurs to some extent in a joint at wide flanges, where screws carry part of the current; parts of the perimeter are likewise in electric contact, but the remaining cross section of the joint acts as a capacitor, having a very thin film of air or corrosion as a dielectric. This capacitor effect is not reliable at low frequencies for obvious reasons, and suitable conductive gaskets usually are needed for wide-frequency shielding, where flanges are not very wide and screw spacings are not very small.

10. Apply by-pass capacitors or low-pass filters (if low frequencies are critical) at points where wires pass through radio-shielded assemblies.

Hints: Filters may be mounted internally or externally; if the latter is the case, short flexible or rigid metal con-

Table II. Relative Copper-Equivalent Thicknesses of Common Metals

Metal	Relative D-C Resistance	Relative C-E Thickness
Silver	0.95	0.975
Copper, annealed, standard	1.00	1.00
Gold	1.42	1.19
Aluminum, pure	1.64	1.29
Aluminum, alloys	1.9–3.2	1.4–1.8
Magnesium, pure	2.67	1.64
Zinc	3.40	1.84
Brass	3.9	2.0
Cadmium	4.4	2.1
Tin	6.7	2.6
Magnesium alloys	7.5–9.8	2.7–3.1
Solder, grade A	8.2	2.9
Lead	12.8	3.6
Nickel (German) silver	16.0–16.9	4.0–4.1
Stainless steel (18 Cr, 8 Ni)	52.8	7.2
Mercury	55.6	7.5
Nichrome	58–65	7.6–8.1
Magnetic metals (iron, steels, nickel and monel, $u=100$ to 10,000)	5–60	0.02–0.8, usually 0.04–0.1

Attenuation $A=3.34t\sqrt{f}/c$ decibels

where t = thickness in mils (0.001 inch); f = frequency in megacycles; and c = relative copper-equivalent thickness.

duits should complete the shielding between the assemblies and the filters, otherwise interference levels may go up instead of down. A filter output terminal or power connection usually is unshielded, but similar terminals sometimes are provided at both ends of the filter to permit reversal in mounting or mechanically shielded power wiring. Either power sources or loads or both may require filters, unless full shielding of the electric system is permissible; usually it is not. As radio interference control and the use of filters approach standardization manufacturers will provide filters within most of their potent products, making for better results at less over-all cost. Meanwhile, many poor shielded leads and unsuitable external filters will be used in critical as well as noncritical installations.

Because of the latter situation, it is well to note that good filters should have from 40- to 70-decibels attenuation at the lowest important frequencies (say, 150 kc) and not less than 40 decibels up to the highest important frequencies (say, 400 megacycles), as tested in the conventional networks, or preferably in service. Some filters contain saturable iron cores, and are not as effective in service as in test. Also, filters should have suitable mounting means, well-shielded housings, at least one shielded terminal, shielded components, if of multiple-type construction, adequate current rating (allowing for temperature rise and voltage drop), adequate voltage rating (allowing for ripple and surge voltages), and resistance to environmental conditions (altitude, corrosion, moisture, temperature, and vibration).

Filter characteristics are interesting, if not exasperating. Many "good" and "bad" resonances occur in the series and shunt elements. The lowest resonant frequency, governed by the nominal values of capacitance and inductance, should be well below the lowest useful radio frequency, in order to maintain adequate attenuation and not gain in the working frequency range.

The next resonant frequency, often around 1 megacycle in commercial filters, is due to self-resonance in the filter coil, and is a "good" frequency, as it gives almost unreadable values of attenuation in some cases, say, 120 decibels. Unfortunately, distributed capacitance in the filter coil has an increasingly detrimental effect at higher frequencies, and series inductance and resistance in the capacitors accent this effect. Few filters have uniformly high attenuation at frequencies from 1 to 400 megacycles. Fortunately, most interference sources fall off rapidly in potency at frequencies above 10 or 100 megacycles, but there are cases where the best available filters are not good enough.

At frequencies from 150 kc to 1 megacycle, the conducted interference voltage at filter output terminals follows an approximate inverse fifth-power law with frequency, due to an approximate inverse-square law, followed by the interference source, and an approximate inverse-cube law, followed by the filter output-to-input voltage ratio. To illustrate the inverse-square law: an inverter delivered 600,000 microvolts at 0.2 megacycle and 6,000 microvolts (roughly) at 2 megacycles. After applying a good filter to the a-c output terminals, the readings were: 50 microvolts at 0.2 megacycle and 2 microvolts at 0.4 megacycle. Thus, the conducted interference voltage dropped 25 to 1 in a 2-to-1 frequency change, or nearly 32 to 1 as given by the inverse-fifth-power law. Similar results were obtained with a different filter applied to the d-c terminals.

Nine years ago Bendix-Scintilla started to develop magneto filters. At first they were single-unit magneto-mounted types; now they are multiple-unit firewall-mounted types. They have been widely used since 1947. Four-unit types are now predominant because of the trend toward low-tension ignition systems having four contact breakers and ground leads, but some filters contain as many as 14 units. Compactness, operating, and grounding requirements limit the electric constants and low-frequency attenuation of magneto filters, but interference seldom occurs in the 200- to 500-kc band and is reducible by using metal-braided cable, an added wire between filter and switch housings, more distance from other wires, or less distance from metallic structure.

Magneto filters serve many purposes such as:

1. to compensate for poorly shielded ground leads;
2. to avoid the cost and weight of ground conduits;
3. to simplify wiring;
4. to save space taken by conduit and elbow bends;
5. to eliminate shielded switches; and
6. to permit the use of starting coils (boosters) and ignition analyzers having imperfect shielding and no power-line filters.

Appendix I. Shielding Theory

The shielding of electric circuits by enveloping metallic conductors is due to skin effect largely; understanding of the latter leads to an understanding of shielding.

Direct current is distributed uniformly in the cross section of a conductor; the current density at any point is simply the total current divided by the total area, as given by the relation $J=I/A$. High-frequency alternating current is not distributed uniformly in metallic conductors; the current density is proportional to the total current but decreases markedly with increasing depth, having a maximum value at the surface and a very low but finite minimum at the center of a solid conductor or at the other surface of a hollow conductor. As most of the current flows in a skin layer, this phenomenon is called "skin effect."

Skin effect reduces the effective cross section, raising the effective resistance from the d-c ohmic value and producing more power loss as heat for any equal rms current and conductor length. The effect increases with frequency, conductivity, and permeability, and varies with the size and shape of conductors. Despite the variable resistance feature, skin effect is an impedance phenomenon. Each imaginary infinitesimal filament in the conductor is encircled by magnetic flux due to its own current and currents in deeper filaments; thus the flux density, the inductance, the reactance, and the impedance increase toward the center of the conductor, producing lower current densities at greater depths. For a cylindrical conductor, $R_{a-c}/R_{d-c}=D/4d+0.25$, where D is the diameter of the conductor and d is the skin depth; accurate when D/d is greater than 4.

The skin effect in a thick hollow conductor is often so complete that two high-frequency currents may flow on opposite surfaces without interaction. An electrically thick wall is virtually divided into three parts: two conducting walls separated by a currentless shield. Thus a metal tube containing an insulated wire forms a self-shielded coaxial-conductor line. The same tube and an external wire may form an unshielded parallel-conductor line. These two lines are shielded from each other, despite their proximity and dependence upon the same intermediate conductor, by virtue of the skin effect. The coaxial line utilizes the inner surface of the tube and the parallel line utilizes the outer surface of the tube; in both cases currents flow in all conductors in such a way that most of the current takes a path of least impedance, forming a loop of minimum area or inductance.

Currents on the outer surface of a coaxial line may be due to external circuits or electromagnetic waves in space; in either case the coaxial circuit is unaffected. Likewise, the coaxial circuit does not affect external circuits and does not produce electromagnetic waves in space. Shielding is bidirectional: it excludes external influences and it confines internal influences. Shielding is a means of decoupling two or more circuits and/or waves without changing either the angular or linear displacement of any circuit component.

A similar discussion would apply to the shielding of parallel-conductor lines. Again, the skin effect is the basis of shielding phenomena, and it matters little whether or not the added shield is grounded, or connected to one of the conductors or to their electric neutral (a center-tapped impedance) at one end or both ends of the line. In all cases, the completeness of the skin effect, or electrical thickness of the shield, largely determines the shielding effectiveness, geometry being of secondary importance. Theoretically a twin-wire line is easier to shield than a coaxial line; practically it is less efficient, more costly, and unnecessary at higher frequencies.

All transmission lines are really wave guides. Their conductors guide wave energy along adjacent dielectric media and act primarily as boundaries, absorbing only enough energy to guide the transmitted energy. The attenuation of twin-wire and coaxial lines increases in proportion to length and approximately as the square root of the frequency, as measured in logarithmic units (decibels or napiers). Transmission efficiency is closely related to the cross-sectional perimeter of the two conductors. At higher frequencies, due to skin effect, it is important to use a type of line which provides conductors of larger perimeter. Thus a coaxial line is superior to a twin-wire line, and a wave guide, having no center conductor, is superior to a coaxial line at applicable frequencies.

An electrically thick shield is one sufficiently thick to greatly attenuate transmitted waves. In such a shield the effect of internal reflection from the remote surface can be neglected. If the wall offers at least 20 decibels attenuation, or 10-to-1 current or field intensity ratio, the reflected wave is subject to at least 40 decibels attenuation during a round trip, even if the remote metal-air boundary is a perfect reflector. As the reflected wave has less than 1 per cent of its initial intensity under this condition, its effect on the incident wave is negligible.

The current density at a point within a thick metal slab decreases exponentially with depth and has a phase lag proportional to depth, as given in the relations

$$j_z/j_0=\epsilon^{-z/d}\epsilon^{-jz/d} \tag{1}$$

$$A=8.686z/d=3.34z\sqrt{fgu}\text{ decibels} \tag{2}$$

$$B=57.3z/d=22.0z\sqrt{fgu}\text{ degrees} \tag{3}$$

where

$$d=2.61/\sqrt{fgu}\text{ mils} \tag{4}$$

f=frequency in megacycles

g=metal/copper conductivity ratio (decimal, except silver=1.05, copper=1.00)

u=metal/air permeability ratio (1 for non-magnetic metals, greater than 1 for magnetic metals)

z=depth in mils (0.001 inch)

The depth d is known as nominal penetration depth or skin depth, at which the current density is attenuated 1 napier ($j_z/j_0=1/\epsilon=1/2.72=0.37$ or 8.686 decibels) and the phase lag is 1 radian (57.3 degrees). If the conductor had a thickness equal to this depth, its d-c resistance would equal the a-c resistance of a much thicker conductor of the same remaining dimensions and metal. The factor z/d, or number of skin depths, measures electrical depth for a given frequency and metal; it also measures attenuation and phase lag.

If desired, the absolute value of the current density ratio (surface/depth values) can be expressed as

$$j_0/j_z=\epsilon^{az}$$

where $a=0.384\sqrt{fgu}$ napiers per mil ($a=1/d$).

Table I is based on these relations and helps to bring out a number of important points, namely

1. One skin depth gives negligible shielding, disregarding reflection effects, which may prove important.
2. Copper having a thickness of 10 mils (0.010 inch) is a good shield, but only at frequencies above 1 megacycle.
3. A minimum frequency of 0.159 megacycle (159 kc) is useful for shielding estimates. It is only 6 per cent above the lowest commonly specified test frequency of 0.150 megacycle (150 kc); it is $1/2\pi$ megacycle and produces a reactance of 1 ohm in either a 1-microhenry coil or a 1-microfarad capacitor; also its square root is very nearly 0.4, simplifying calculations. Thus, for copper at 0.159 megacycle, $A=1.33\ z$ or $z=0.75\ A$ mils.
4. The boundary between electrically thin and electrically thick shielding may be assumed as 2.3 skin depths or 20 decibels, above which the relations just given are quite accurate, and below which special formulas including reflection effects must be used. However, the mathematical complexity of the electrically thin case seldom is warranted, because better shielding normally is available and required.
5. Wave length in metals is only 6.28 skin depths, within practical limits of thickness, and varies inversely as the square root of frequency, not inversely as in space.
6. Table I makes no allowance for shielding defects, such as discontinuities, holes, shafts, and slots in the joints and walls of shielding assemblies.
7. To obtain at least 40 decibels (100 to 1) attenuation at frequencies above 0.159 megacycles, 30 mils thickness of copper or equivalent is required. The copper-equivalent thickness of another metal is independent of frequency, namely

$$z^1=z/\sqrt{gu}$$

8. Thus, for brass, $g=0.25$ roughly, and $z^1=2z$, so that twice the copper thickness or 60 mils thickness is required. For other non-magnetic metals the thickness ratio ranges from 1.3 (pure aluminum) to about 7.5 (some nickel-steels) times the thickness of copper. See Table II.

References

1. NOTES ON INFORMAL CONFERENCE RELATIVE TO INCIDENTAL RADIATION DEVICES HELD AT WASHINGTON, D. C., NOVEMBER 1, 1949. *Mimeograph #43869, 12–49, Docket 9288*, Federal Communications Commission, Washington, D. C., 1949.

2. STANDARDS ON RADIO RECEIVERS. Institute of Radio Engineers (New York, N. Y.), 1938.

3. METHODS OF MEASURING RADIO NOISE, Joint **EEI-NEMA-RMA** Report. *EEI Number G9, NEMA Number 107, RMA Number 32* (New York, N. Y.), February 1940.

4. METHOD OF MEASURING RADIO INTERFERENCE OF ELECTRICAL COMPONENTS AND COMPLETED ASSEMBLIES OF ELECTRICAL EQUIPMENT FOR THE ARMED FORCES FROM 150 KILOCYCLES TO 20 MEGACYCLES. *ASA Number C63.1-1946*, American Standards Association (New York, N. Y.), 1946.

5. SPECIFICATIONS FOR RADIO NOISE METER, 0.015 TO 25 MEGACYCLES/SECOND. *ASA Number C63.2-1950*, American Standards Association, *NEMA Number 102-1950, RMA Number 32-A* (New York, N. Y.), 1950.

6. 1952 SAE HANDBOOK. Society of Automotive Engineers, New York, N. Y., 1952, page 763.

7. CYLINDRICAL SHIELDING, **A. R. Anderson.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 34, 1946, pages 312–22.

8. RADIO SHIELDING OF AIRCRAFT ENGINE IGNITION SYSTEMS. *TN-92-1*, United States Air Force (Dayton, Ohio), 1946.

9. ENGINE-IGNITION SHIELDING FOR RADIO RECEPTION IN AIRCRAFT, **H. Diamond, F. G. Gardner.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 18, 1930, pages 840–61.

10. FACTORS AFFECTING THE ACCURACY OF RADIO NOISE METERS, **F. E. Dinger, H. G. Paine.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 35, 1947, pages 75–81.

11. REFERENCE DATA FOR RADIO ENGINEERS (book). Federal Telegraph and Radio Company, New York, N. Y., third edition, 1949, pages 296–303.

12. RADIO-NOISE ELIMINATION IN ALL-METAL AIRCRAFT, **Fred Foulon.** *AIEE Transactions*, volume 62, 1943, pages 877–91.

13. A STUDY OF WAVE SHAPES FOR RADIO-NOISE-METER CALIBRATIONS, **C. W. Frick.** *AIEE Transactions*, volume 64, 1945, pages 890–901, discussion 969–70.

14. RADIO-NOISE FILTERS APPLIED TO AIRCRAFT, **C. W. Frick, S. W. Zimmerman.** *AIEE Transactions* (*Electrical Engineering*), volume 62, September 1943, pages 590–95.

15. NOISE LEVELS IN THE AMERICAN SUB-ARCTIC, **N. C. Gerson.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 38, 1950, pages 905–08.

16. FREQUENCY ANALYSIS, MODULATION AND NOISE (book), **S. Goldman.** McGraw-Hill Book Company, New York, N. Y., 1947.

17. NOISE REDUCTION AND RANGE IN RADAR AND COMMUNICATIONS, **S. Goldman.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 36, 1948, pages 584–94.

18. THE INFLUENCE OF THE GROUND ON THE CALIBRATION AND USE OF VHF FIELD INTENSITY METERS, **F. M. Green.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 38, 1950, page 650.

19. MEASUREMENTS OF RECEIVER IMPULSE NOISE SUSCEPTIBILITY. Measurements Notes, Measurements Corporation (Boonton, N. J.), number 2, 1950.

20. ELIMINATION OF RADIO INTERFERENCE IN AIRCRAFT. *Nav Aer 16-5Q-517* (restricted), 1946. Abridged issue *PB81895* (unclassified). Bureau of Aeronautics, United States Navy Department (Washington, D. C.)

21. RADIO RECEIVER IGNITION INTERFERENCE TESTER, **W. D. Nupp.** United States Patent *2,499,410*.

22. TRACKING RADIO NOISE ON AIRCRAFT, **T. B. Owen.** *AIEE Transactions*, volume 69, part II, 1950, pages 1249–52.

23. VERY HIGH-FREQUENCY RADIO-NOISE ELIMINATION, **T. B. Owen.** *AIEE Transactions* (*Electrical Engineering*), volume 63, December 1944, pages 949–54.

24. SPECIFICATION AND MEASUREMENT OF RECEIVER SENSITIVITY AT THE HIGHER FREQUENCIES, **J. M. Pettit.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 35, 1947, pages 302–06.

25. EFFECTIVENESS OF CONDUIT AS R-F SHIELDING, **S. L. Shive.** *Electronics* (New York, N. Y.), volume 19, February 1946, pages 160–66.

26. RADIO NOISE IN AIRCRAFT ENGINES, **A. E. Teachman.** *Aeronautical Engineering Review* (New York, N. Y.), volume 4, August 1945, pages 13–32.

27. RADIO-NOISE ELIMINATION IN MILITARY AIRCRAFT, **G. Weinstein, H. H. Howell, G. P. Lowe, B. J. Winter.** *AIEE Transactions* (*Electrical Engineering*), volume 63, November 1944, pages 793–95.

28. TELEVISION INTERFERENCE. Remington Rand Laboratory of Advanced Research, South Norwalk, Conn. (Reprints of articles which have appeared in *CQ, Electrical Manufacturing, Electronics, Modern Plastics, Radio and Television News, QST*, March

1946 to October 1951.) (Second edition includes late 1951 articles, other periodicals.)

29. SHIELDING FOR DIATHERMY, W. C. Stokes, W. W. Seifert. *Docket Number 6651*, Federal Communications Commission, Washington, D. C.

30. SHIELDING HIGH FREQUENCY INTERFERENCE, A. F. Murray. *Electronic Industries* (New York, N. Y.), volume 4, August 1945.

31. ELECTROMAGNETIC WAVES (book), S. A. Schelkunoff. D. Van Nostrand Company, Inc., New York, N. Y., 1943, pages 303–15.

32. SKIN EFFECT FORMULAS, J. R. Whinnery. *Electronics* (New York, N. Y.), volume 15, February 1942, pages 44–48.

33. RADIO ENGINEERS HANDBOOK (book), F. E. Terman. McGraw-Hill Book Company, New York, N. Y., 1943, pages 770–72, 981–82.

34. ELECTROMAGNETIC THEORY (book), J. A. Stratton. McGraw-Hill Book Company, New York, N. Y., 1941.

35. TRANSMISSION NETWORKS AND WAVE FILTERS, T. E. Shea. D. Van Nostrand Company, Inc., New York, N. Y., 1929, pages 377–445.

36. RADIO-FREQUENCY DISTRIBUTING SYSTEMS, F. X. Rettenmeyer. *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 23, 1935, pages 1286–1307.

37. CAR IGNITION RADIATION, C. C. Eaglesfield. *Wireless Engineer* (London, England), volume 28 number 328, January 1951, pages 17–22.

No Discussion

A New General-Purpose Relay for Telephone Switching Systems

A. C. KELLER
MEMBER AIEE

THIS paper describes a new general-purpose electromagnetic relay for use in telephone switching systems. It is a wire spring relay known as the *AF*-type relay and, with variations which provide slow release or marginal characteristics, it is known as the *AG* and *AJ* relay respectively. Figure 1 shows a typical *AF*-type relay, Figure 2 shows all of the parts of the relay, and Figure 3 shows the relay assembly.

Background

The general-purpose relay is one of the most important components of telephone switching systems.[1] These relays constitute the most repetitive building block in switching equipment. Since several million are produced annually, low manufacturing cost is extremely desirable. Also of prime importance are low operating and maintenance costs. General-purpose relays are therefore under constant observation and study by the telephone operating companies as the users, by the Western Electric Company as the manufacturer, and by Bell Telephone Laboratories as the designer. The *AF* wire spring relay and its variations are the result of such studies.

A general-purpose relay for telephone switching systems must meet a large number of diverse requirements. It must be capable of being assembled with any one of a variety of magnet coils having a wide range of resistance values and of operating contacts which vary from one pair to as many as 14 or more. The basic relay design also must be capable of providing such features as fast operation and release, slow release, high sensitivity, heavy duty, and marginal operation. These functions are performed satisfactorily in present cross-bar switching systems by *U*-, *UA*-, *UB*-, and *Y*-type relays.[2-5] However, with an objective of a 40-year life for new switching systems and a trend toward unattended operation of switching offices, it is important to attain the best in the performance and reliability of relays.

The general-purpose relay must be designed to produce the best economic balance, when used in telephone switching systems, so that the annual charges are minimized. The major ingredients of these annual charges are manufacturing expense, operating electric power, speed of operation and release, space required, and maintenance costs which include reliability and life.

Requirements and Objectives

The requirements for a new general-purpose relay were initially broadly stated to be performance and maintenance at least equal to the *U*- and *Y*-type relays but with substantially lower manufacturing costs. As the development of the relay proceeded, it became possible to expand the requirements without appreciably altering the expected relay cost. In particular, it became possible to design the new relay to operate and release faster or to use less electric power, to operate more often before appreciable wear occurred, and so forth. The improved performance characteristics of the *AF* wire spring relay, as described later in this paper, are of equal economic importance to those associated with lower manufacturing cost.

The broad requirements were reduced to the following design objectives:

1. Lower cost—50 per cent of *U*-type relay.
2. Reduced operating electric power.
3. Faster operate and release times.
4. Long life—one billion operations.
5. Improved contact performance.

These broad design objectives do not specifically state a large number of other characteristics which must be at least as favorable as those of the *U* and *Y* relay family. This refers to such items as: space required, magnetic interference, wiring costs, contact combinations, field servicing, and repairs.

Design Possibilities

The studies of new relay design possibilities started with a careful review of the *U*-type relay experience. In fact, much of the early thinking considered various modifications of *U*-type and other existing relays. In general, these studies indicated that about half of the manufacturing cost of *U*-type relays came from

---

Paper 52-304, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted July 14, 1952; made available for printing September 4, 1952.

A. C. KELLER is with Bell Telephone Laboratories, Inc., New York, N. Y.

The development of the new relay described in this paper has called for a major co-operative effort of many sections of Bell Telephone Laboratories, including such departments as Switching Apparatus Development, Switching Systems Engineering, Switching Systems Development, Research, Materials, and Chemical. Without the co-operation of the many members of these organizations and their special skills, this development would not have resulted in the important and balanced design which has been described. Throughout the development, the associated organizations in the Western Electric Company and the American Telephone and Telegraph Company have made important and guiding contributions, and the New York Telephone Company has co-operated in field trials.

I wish to express appreciation to the many people who have contributed to this important technical accomplishment. The few names which have been mentioned in the paper were given for historical reasons and cannot replace the large number of important contributors to the project. To D. C. Koehler, I am indebted for assistance in preparing the paper and the illustrations.

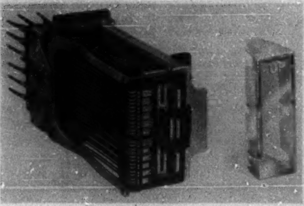


Figure 1 (left). AF-type relay, with contact cover detached

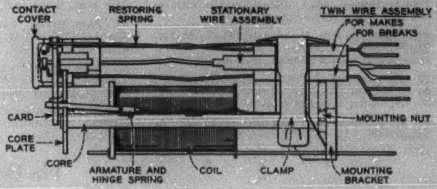

CONTACT COVER
RESTORING SPRING
STATIONARY WIRE ASSEMBLY
TWIN WIRE ASSEMBLY
FOR MAKES
FOR BREAKS
CARD
MOUNTING NUT
CORE PLATE
CORE
ARMATURE AND HINGE SPRING
COIL
CLAMP
MOUNTING BRACKET

Figure 3 (right). Top view of the relay, showing location of parts

assembly and adjusting operations. Accordingly, these operations required major revision for a substantial cost reduction of the relay. It became evident that the development of new manufacturing methods as well as new designs were essential in reaching the ambitious objectives. For these reasons, the manufacturing engineers of the Western Electric Company were active participants in the development of the new relay from the beginning.

Many new forms of relay designs were considered and studied including such types as miniature, magnetic contact, piezoelectric, and so forth. As a result, one general form, first proposed by H. C. Harrison, gave the most promise of meeting the manifold requirements. This is the wire spring type characterized by the wire spring subassemblies with code card operation of pretensioned, low-stiffness springs. Actually, the general form of the wire spring relay proposed by Mr. Harrison constitutes an entire new class of relays with many possible variations. These include various types of code card operation and various forms of contact operation, operated by any of a number of magnet structures.

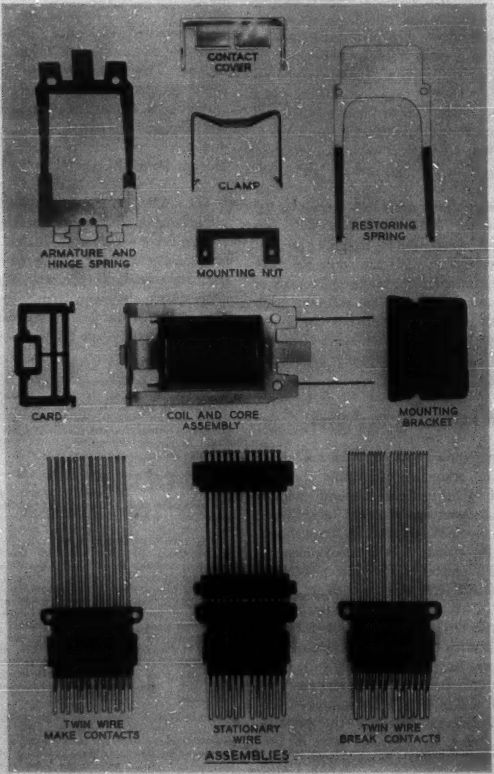

CONTACT COVER
ARMATURE AND HINGE SPRING
CLAMP
RESTORING SPRING
MOUNTING NUT
CARD
COIL AND CORE ASSEMBLY
MOUNTING BRACKET
TWIN WIRE MAKE CONTACTS
STATIONARY WIRE
TWIN WIRE BREAK CONTACTS
ASSEMBLIES

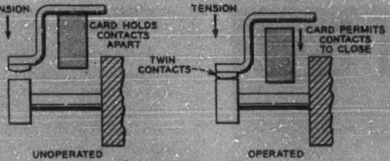
Figure 2. Parts of the wire spring relay

The new class of relays has the following important advantages:

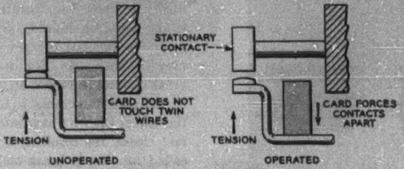
1. Pretensioned, low-stiffness wire springs make possible
   - a. assembly to give close control of contact force without individual spring adjustment;
   - b. essentially constant contact force throughout the life of the relay and its contacts.
2. Wire spring subassemblies make possible
   - a. favorable manufacture of a multiplicity of contact springs by molding;
   - b. lower assembly costs because fewer piece parts are needed;
   - c. simple code card operation.
3. Code card operation makes possible
   - a. standardized and simple assembly;
   - b. accurate control of contact position;
   - c. essential elimination of locked contacts;
   - d. complete independance of twin contacts;
   - e. simple means for providing a large number of contact combinations.

A continuous and comprehensive study of the characteristics and probable manufacturing costs of many forms of the wire spring relay family was necessary. As a result, after passing through several major designs, the basic design of the present relay was adopted. H. M. Knapp and C. F. Spahn proposed important features of this design. This form represented advantages over other types in

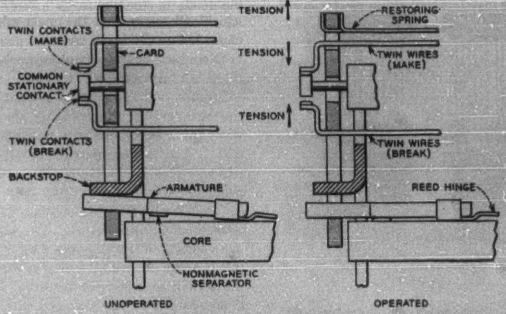
1. reducing the number and amount of dimensional variations controlling the contact gaps. In turn, this made possible smaller armature movement, shorter operating and release times, and less chatter of the contacts;
2. reducing the number of code cards required to provide the large number of contact combinations needed in switching systems;


TENSION
CARD HOLDS CONTACTS APART
TENSION
CARD PERMITS CONTACTS TO CLOSE
TWIN CONTACTS
UNOPERATED
OPERATED

(a) MAKE CONTACT


STATIONARY CONTACT
CARD DOES NOT TOUCH TWIN WIRES
CARD FORCES CONTACTS APART
TENSION
TENSION
UNOPERATED
OPERATED

(b) BREAK CONTACT


TENSION
RESTORING SPRING
TWIN CONTACTS (MAKE)
CARD
TENSION
TWIN WIRES (MAKE)
COMMON STATIONARY CONTACT
TENSION
TWIN CONTACTS (BREAK)
TWIN WIRES (BREAK)
BACKSTOP
ARMATURE
REED HINGE
CORE
NONMAGNETIC SEPARATOR
UNOPERATED
OPERATED

(c) TRANSFER CONTACT

Figure 4. Principle of contact operation

3. reducing the manufacturing and wiring costs;
4. increasing the mechanical life.

Principle of Contact Operation of the *AF* Relay

The *AF* relay uses what has been called the "single card system" for actuating the contacts. This is in contrast to other code card systems which require 2, 3, or 4 coded cards in each relay. The method for obtaining individual make and break contacts with this system is shown in Figures 4(A) and 4(B), and a means for obtaining transfer contacts, in which both make brand eak twin contacts are associated with a common stationary contact, is shown in Figure 4(C). As indicated in Figure 4, the following principles are incorporated in this method of actuation:

1. In general, three basic wire spring assemblies are required. Two of these carry movable twin wires for make and break contacts and are identical except for some details in forming at the terminal ends for convenience in wiring. The twin wire assemblies are mounted on either side of the stationary wire assembly, which consists of a group of relatively heavy wires molded into plastic sections, one a short distance behind the contacts and one near the rear of the relay. These sections are rigidly supported in the relay structure.

2. Moving twin contacts on separate twin wires are used with every stationary contact. This arrangement assures good reliability and greater freedom from open contacts in the presence of dust and dirt. In addition, contact chatter is reduced as both contacts must be open simultaneously in order to interrupt the circuit.

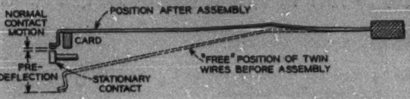
3. As shown in Figure 5, each group of twin wires is tensioned toward the stationary wires by means of large predeflections before assembly, so the contact forces are determined by this predeflection. Good control of the contact force is assured without need for hand adjustment because small variations in deflection of the low-stiffness springs do not result in appreciable changes in force. For this reason, the force is stable and is not appreciably affected by wear of the contacts.

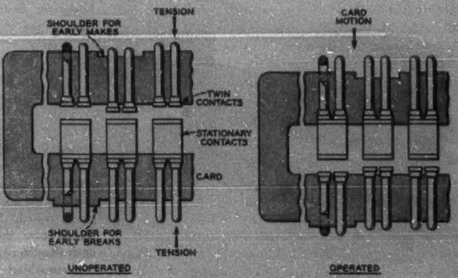
4. The twin wires are actuated by a single punched fiber card. Since the tension in the twin wires is always in a direction to hold the contacts closed, the card serves to hold the make contacts open when the relay is unoperated and the break contacts open when the relay is energized.

5. The card is supported by the armature on one side and a restoring spring on the other. The restoring spring supplies the force to hold the armature against the backstop and to hold make contacts open when the relay is unoperated, while the armature supplies the force to hold the break contacts open when the relay is operated. However, since the armature also must overcome the tension in the restoring spring, the entire spring load of course must be overcome by the pull of the armature.

Figure 5 (below). Contact forces are controlled by relatively large predeflections of the twin wires

Figure 6 (right). Early break-make, early make-break, and nonsequence transfer contacts, showing how early contacts are obtained by means of shoulders on the actuating card


NORMAL CONTACT MOTION
POSITION AFTER ASSEMBLY
CARD
PRE-DEFLECTION
STATIONARY CONTACT
"FREE" POSITION OF TWIN WIRES BEFORE ASSEMBLY


SHOULDER FOR EARLY MAKES
TENSION
CARD MOTION
TWIN CONTACTS
STATIONARY CONTACTS
CARD
SHOULDER FOR EARLY BREAKS
TENSION
UNOPERATED
OPERATED

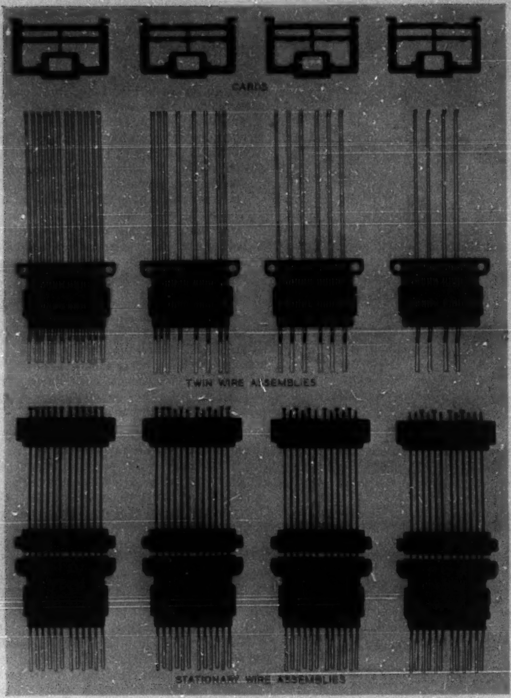


Figure 7. A few varieties of the coded parts used to obtain various contact combinations

6. The twin contacts are held in good registration with their associated stationary contacts by means of molded guide slots in the stationary plastic member just behind the card. These guide slots are slightly wider than the diameter of the twin wires so that these wires are free to move in the direction of the armature movement, but are restrained against lateral motion.

7. The close proximity of the card to the contacts is important in minimizing contact chatter and in substantially eliminating locked contacts, that is, contacts which fail to open because of interlocking of roughened surfaces. The close spacing results in a rigid coupling between the card and contacts, so that the static and dynamic forces associated with the armature and card are available to break loose any incipient lock which might develop.

As the armature moves toward the core, the particular point in its travel at which make contacts close and break contacts open depends upon the dimensions of the card between the surface which bears against the armature and the surfaces which engage the twin wires. By proper selection of these dimensions, any contact can be controlled to operate early or late in the travel as desired. By this means, several sequential contact arrangements may be obtained. For example, if the break contact in Figure 4(C) is controlled by the card dimensions to open earlier in the travel than its associated make contact closes, the resulting arrangement is called an "early break-make" transfer. Similarly, an "early make-break" transfer, often called a "continuity," may be obtained by selection of card dimensions which will assure that the make contact closes before the break contact opens. If both contacts operate simultaneously, the result is a "nonsequence" transfer.

From the preceding it is evident that the card surfaces which engage the twin wires must be in different positions for early contacts as compared with late contacts. This is illustrated in Figure 6 which shows an early break-make, an early make-break, and a nonsequence transfer side by side. Of the contact pairs shown, only two operate early, and this is accomplished by means of steps in the actuating surfaces of the card. Thus, if no sequences were required, the card would have a single straight surface for makes and another for breaks, and only one card variety would be needed for all combinations of makes, breaks, and nonsequence transfers. Where sequences are needed, however, additional card varieties are required with steps in the actuating surfaces for the early contacts.

To obtain a wider variety of the contact combinations including various numbers of make contacts, break contacts, sequence transfers, and nonsequence trans-

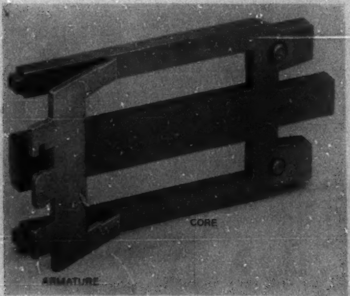


Figure 8 (left). Magnetic structure of the AF relay

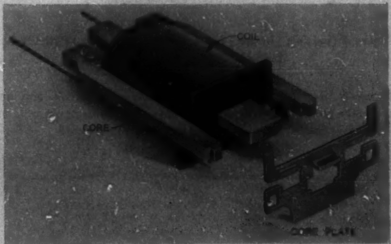


Figure 9 (right). Legs of the core are held in alignment by the core plate, which is forced over the ends after the coil is assembled

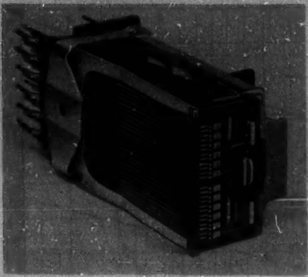


Figure 10. AF relay with 12 make contacts

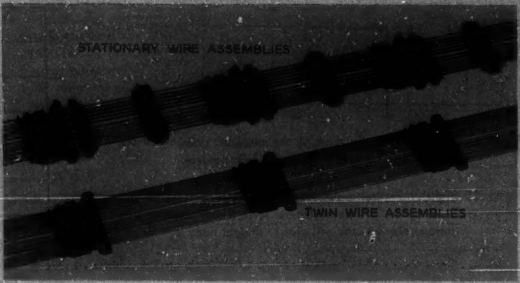


Figure 12. Molded stationary and twin wire assemblies, before cutting to length

fers on the same relay, it is necessary to provide a variety of different coded stationary and twin wire assemblies, as well as a variety of cards, some of which are illustrated in Figure 7. The twin wire assemblies differ as to the number of twin wires provided and in the position of these wires across the width of the molded section.

The stationary wire assemblies always are provided with a full complement of 12 wires in order to support the front molded section, which is held in place by spring tension in these wires. However, only certain of the wires may have contacts at the ends. These stationary contacts consist of base metal blocks with 0.010-inch-thick precious metal surfaces on either or both sides as needed for makes, breaks, or transfers, and any of the three varieties may be welded to any wire. Thus precious metal is provided only where needed for the particular contact arrangements desired.

By using different combinations of stationary and twin wire assemblies with each card variety, a large number of different contact combinations may be obtained. While most of these needed for telephone switching systems use either no sequences at all or a single stage of sequence, a few combinations are provided with "preliminary" contacts. These combinations include two stages of sequence, in which some contacts operate at each of three different points in the armature travel. The preliminary contacts operate earliest in the travel. These are followed by the early contacts of sequence transfers and finally by the late contacts, including ordinary makes and breaks.

To be sure the desired sequences will be maintained during the life of the relays, it is necessary to provide margins in the form of armature travel allowances at each stage. Combinations with sequences therefore will require total armature travels which are longer than those with no sequences, and two stages of sequence will require more travel than a single stage. Accordingly, the *AF* relay is provided with a choice of three armature travels to correspond with the number of sequences needed. At the card, these travels are 0.026 inch (short) for no sequences, 0.044 inch (intermediate) for one

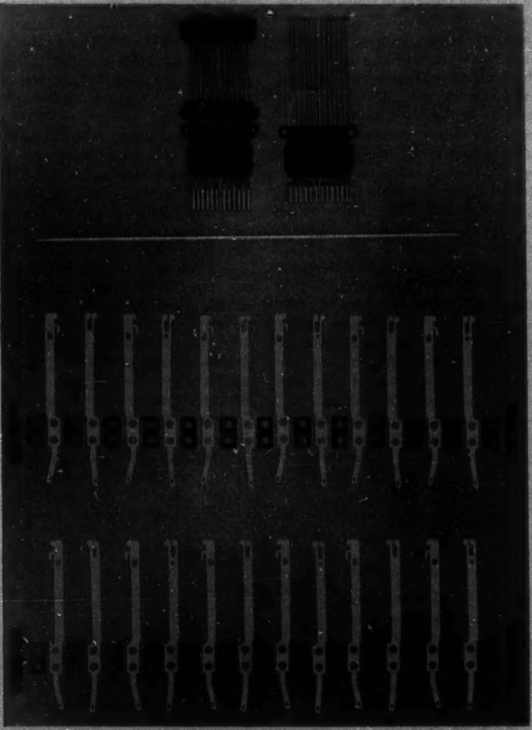


Figure 11. A comparison of the molded wire assemblies for 12 make contacts with the corresponding U relay parts

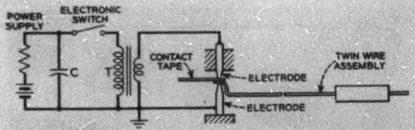
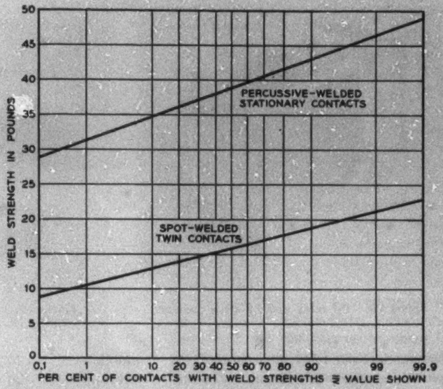


Figure 13 (above). Diagram showing the essential elements of the spot welding process used for the twin contacts

Figure 15 (right). Typical weld strength distributions for the stationary and twin contacts. The horizontal scale is graduated so that normal distributions will plot as straight lines



stage, and 0.060 inch (long) for two stages.

Thus, combinations including ordinary makes, breaks, and nonsequence transfers use short travel. Where sequence transfers also are needed, intermediate travel is used and the early contacts of the sequence transfers operate first. Long travel is used only where preliminary contacts followed by sequence transfers are needed.

Armature System and Magnetic Circuit

The armature system and the associated magnetic circuits constitute the basic motor element of an electromagnetic relay. The size of the motor element is determined, in part, by the work it must do and here a basic factor is the contact force. On the basis of analytical as well as experimental studies, it was decided to use a contact force of about 6 grams per single contact, that is, about 12 grams for the combined force of the twin contacts Other important factors which react on the design of the magnet are the speed required, winding space, heating,[6] sensitivity, and so forth. The detailed analysis of the magnetic system and the associated measurements will be covered in separate papers.

The magnetic structure chosen is shown in Figure 8. The armature is a flat member of U shape which provides desirably large pole-face areas. The core is a simple 1-piece E-shape section of sufficient thickness of 1-per-cent silicon iron to produce the magnetic flux needed to meet the force and speed requirements and to provide the main member to which all other parts are assembled. The silicon iron has appreciably higher electrical resistance than ordinary magnetic iron and this, together with the rectangular cross sections of the legs, reduces eddy currents as needed for high-speed operation and release. The 1-piece construction avoids welded or butt joints common to many magnets. These joints are responsible for added reluctance and hence decrease the magnet sensitivity and require added electric power to operate a given load. The relatively wide spacing of the legs increases leakage reluctance and, in turn, increases the useful magnetic flux.

After a cellulose acetate filled coil[7] has been assembled to the middle leg of the core, a core plate, shown in Figure 9, is forced over the ends of the E-shaped core to hold the three legs in good alignment for proper mating with the armature. The core plate also provides the backstop for the armature and serves as a means of gang adjustment of the contacts covered more completely under the "Relay Adjustment" section of this paper.

The armature is spring-supported in a very definite manner to produce a minimum of rebound when it is released from its operated position. The conditions for reducing armature rebound are described in detail elsewhere[8] and make it necessary to proportion the forces at the front and rear of the armature properly. The magnitude and the ratio of these forces are a function of the mass distribution of the armature.

The magnet design must not only meet such functional requirements as speed, sensitivity, and so forth, but it must meet these for several values of armature travel as needed by the variety of contact combinations provided. Another requirement is that the relay be designed to fit on a 2-inch mounting plate and this, in turn, restricts the width of the E-shaped magnet core to slightly less than 2 inches. Normally the relay is mounted with the 2-inch dimension in the vertical direction to allow the contact surfaces to be in vertical planes. The corresponding horizontal dimension in which the relay can be mounted is $1^{1}/_{2}$ inches except for a few special cases. As described in more detail under the section on "Relay Performance," the improved magnet design has resulted in a reduction of the magnetic interference between mounted relays to values which

Figure 14. Diagram showing the essential elements of the percussive welding process used for the stationary contacts

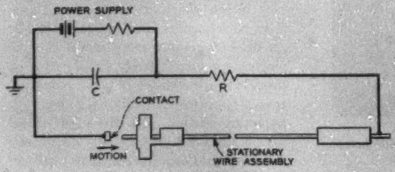
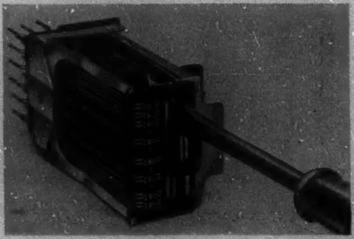


Figure 16 (right). Contacts may be gang-adjusted to operate at the proper points in the armature travel by bending the arms of the core plate with a screwdriver



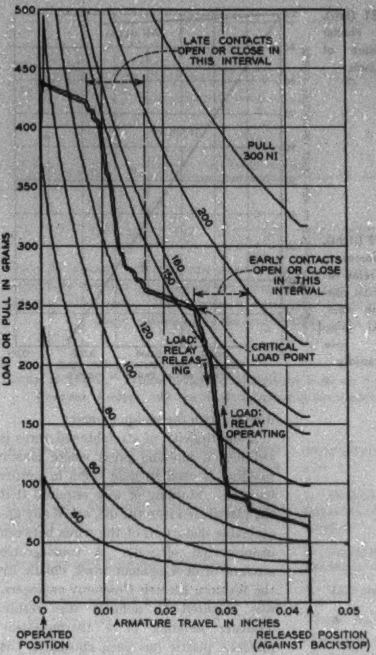


Figure 17 (left). Typical load and pull characteristics of a wire spring relay with 12 early break-make transfer contacts

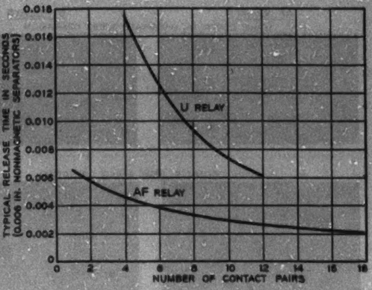


Figure 19 (right). Typical release times of AF and U relays

are negligible for most practical purposes.

For comparison with the *U*-type relay, the following typical constants of the magnet are of interest. The closed gap reluctance $\mathcal{R}_0$ is the reluctance of the magnetic circuit, excluding leakage paths, with the armature operated and with the iron near maximum permeability. The coil constant *G* is the ratio of the square of the number of turns to the resistance for a full-sized coil. The sensitivity *S* is a measure of the ultimate work capacity of the magnet as related to the power input and has been defined as $S=5\pi G/\mathcal{R}_0$ ergs per watt.

The favorable low value of closed gap reluctance for the new relay results from adequate cross sections of magnetic material, the absence of joints, proper mating of the armature and core, and large pole-face areas. A low value of reluctance also insures less magnetic interference to other relays and from other relays.

Although the coil constants are the same for the two relays, as can be seen from Table I, the sensitivity of the new relay is more than double that of the *U* relay because of the lower closed gap reluctance.

Molded Wire Spring Subassemblies

One of the major features of the new relay is the use of molded wire spring subassemblies. Figure 10 shows a wire spring relay with 12 make contacts, and Figure 11 shows a comparison of the wire spring assemblies used in this relay and the corresponding parts of the *U* relay. From this it is clear that the number of parts handled in the assembly of the contact

Table I

	AF Relay	U Relay
Closed gap reluctance $\mathcal{R}_0$, centimeter^{-1}	0.028	0.065
Coil constant *G*, kilomhos	160	160
Sensitivity *S*, ergs per watt	90,000	39,000

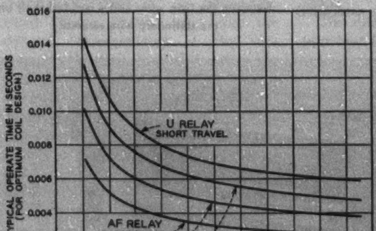


Figure 18 (left). Typical operate times of speed relays with optimum coil designs for high-speed operation

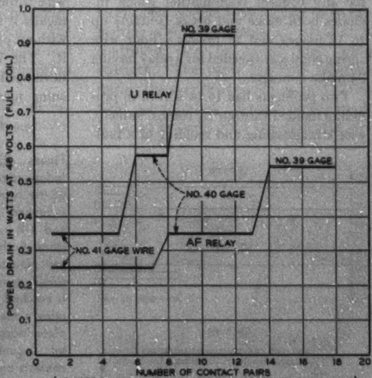


Figure 20 (right). Power used for AF and U relays with coils designed for least power

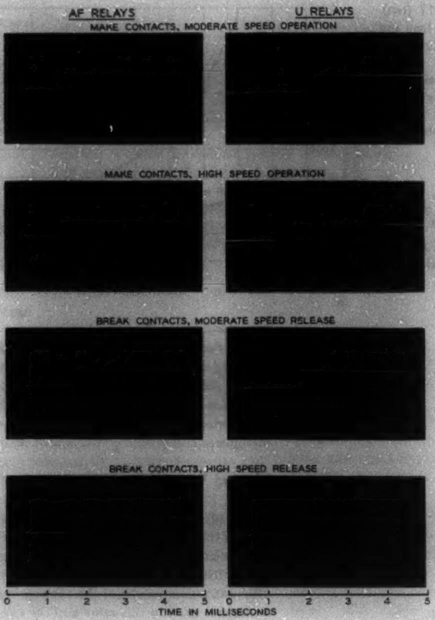


Figure 21 (left). Typical chatter on closure of contacts

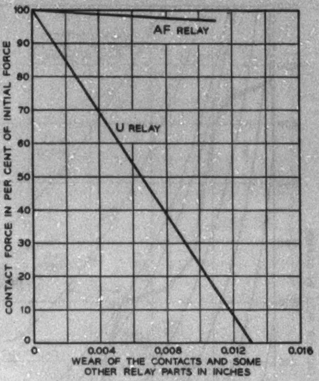


Figure 23 (right). Contact forces on the AF relay remain almost constant with wear, while U relay contacts lose force rapidly

spring members is reduced greatly in the new relay. Not all relays will have 12 contacts, and in those cases where fewer contact springs are needed the comparison will not be so unfavorable to the *U* relay. For six contacts, about one-half of the parts shown will be needed for the *U* relay, whereas the new relay again will require two wire spring combs. In the new relay three wire spring combs are needed for any contact combination which includes both make and break contacts up to 12 makes and 12 breaks. Four wire spring combs are needed for a relay having 24 make contacts.

Two problems had to be solved in providing molded wire spring combs, namely, wire straightening and molding of a multiplicity of wires. Both of these were studied co-operatively at Bell Telephone Laboratories and the Western Electric Company.

Wire is straightened by rotating cam and die members around the unstraightened wire which causes alternating flexing of the wire. For best results, it was found important to shape the cams properly and to push, rather than to pull, the wire through the rotating cams. By this means it is expected to get straight wire without producing an appreciable twist in it. The Western Electric Company has developed a multiple-head wire straightening machine which can be associated directly with the molding press.

A multiplicity of straightened wires is fed into a molding press where plastic molding is used to hold them in proper location. Molding of wire required that the plastic, fed into the die, avoid any appreciable distortion of the wires between unsupported sections. A considerable amount of development work, chiefly by the Western Electric Company engineers, was required to achieve this result. Transfer molding of a thermosetting phenolic plastic has been chosen as the most suitable for producing stable wire spring subassemblies. This is based on the need for stability of the wire positions and on the ability of the material to withstand the effects of heat. Figure 12 shows continuous ladders of molded wire spring sections before cutting to length.

The molded sections have a number of features of design importance beyond holding the wires in place. These added features are provided by shaping molded sections to make the remainder of the re-

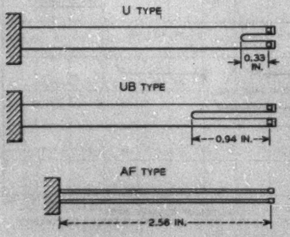
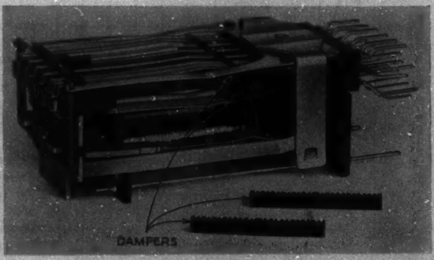


Figure 22 (left). Independent action of the twin contacts is limited on U and and UB relays because both contacts are mounted on a single spring which is notched. The AF relay achieves complete independence by mounting the contacts on separate wires

Figure 24 (below). Where very long life is needed, polyisobutylene dampers are mounted between the twin wires and a molded shelf on the stationary wire assembly



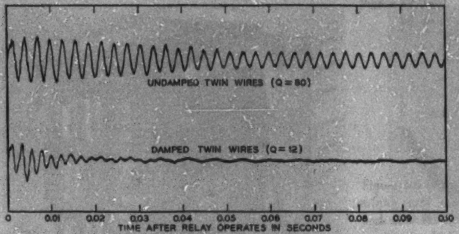


Figure 25 (above). Oscillograms showing the effectiveness of the polyisobutylene dampers in reducing the vibration of twin wires following operation of the relay. The vibration is measured in the horizontal plane, about midway along the lengths of the wires

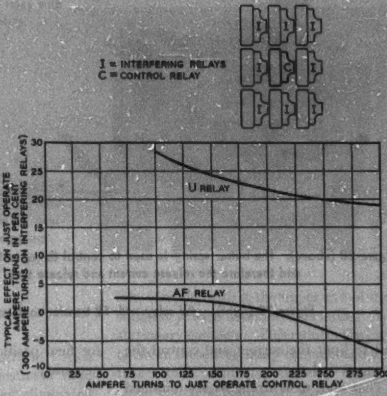


Figure 26 (right). Typical magnetic interference between AF relays and between U relays, with the relays mounted in the pattern shown

lay simpler. In particular, these features provide registration pins and holes, guides for the ends of the twin wires, cover anchorage, damping material support, and so forth.

Contacts and Contact Welding

Since the primary purpose of the relay is to open and close electric circuits through the contacts, there has been a special effort to make these contacts as reliable as possible. Accordingly, palladium is used for all contact surfaces. This use of precious metal substantially eliminates openings due to corrosion. Palladium not only gives outstanding reliability but studies indicate that its use results in the best economic balance between manufacturing cost and service because of the reduced maintenance expense.

Open circuits resulting from particles of dirt between the contact surfaces are eliminated largely by the use of a contact cover, complete independence of the twin contacts described in the section entitled "Relay Performance," and the dynamic characteristics of the wire springs. However, to further reduce the incidence of dirt troubles, the surfaces of the twin contacts are coined to a cylindrical shape. This greatly reduces the effective bearing area between the twin contacts and the flat surfaces of the single contacts. Thus, even if an occasional dirt particle should come to rest on one of the contact surfaces, there is small likelihood that it would be in the contact area.

Since welding contacts to wires instead of flat springs is relatively new, considerable attention was given to the development of suitable techniques. The basic requirements for satisfactory welds are:

1. Sufficient strength to withstand the forces encountered during manufacture and service.
2. Accurate positioning of the contacts on the wires.
3. Low cost.

These requirements apply to both stationary and twin contacts. However, because of differences in geometry, entirely different methods have been developed for welding the two types of contacts.

The twin contacts are produced by spot-welding precious metal contact tapes to the tips of the twin wires. The diagram of the welding circuit is shown in Figure 13. Capacitor *C* is charged by a power supply to a predetermined voltage. The capacitor then is discharged through the primary of the welding transformer *T*, giving rise to the low-voltage high-current surge which produces the weld. The contacts then are sheared to length and the surfaces are coined to a cylindrical shape.

The spot-welding process did not appear best for welding the stationary contacts to the ends of the wires because of the need to grip the wires with heavy welding electrodes in the limited space directly behind the contacts. Accordingly, a type of welding, known as "percussive welding," was developed which permits one of the electrodes to be placed near the wiring end of the wire springs without developing excessive heat in the wires and which also permits the accurate positioning needed for the contacts in order to control the point of contact closure on the assembled relay. The welding circuit is shown in Figure 14. Capacitor *C* is charged by means of a d-c power supply, and the capacitor voltage also appears on the stationary wire. As the contact to be welded is moved toward the end of the wire, the capacitor discharges, forming an arc which melts the abutting surfaces of the contact and wire. The constants of the electric circuit and mechanical system were chosen to assure melting a proper amount of metal at a controlled rate to assure high weld strength. The parts are held together during the very brief cooling period as the weld is completed. A small resistance *R* is added in series with the discharge cir-

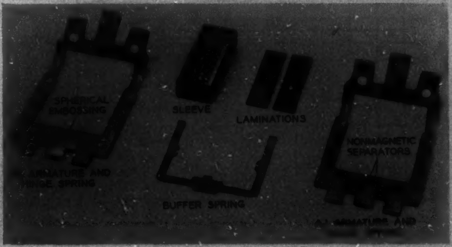


Figure 27. Additional parts for AG and AJ relays

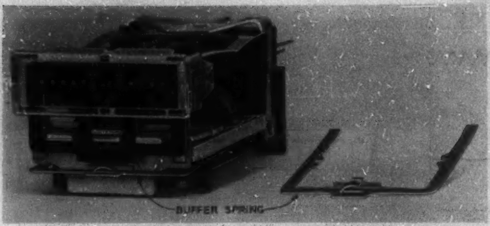


Figure 28 (above). The buffer spring is used to control the operated spring load, and therefore the release current and release time

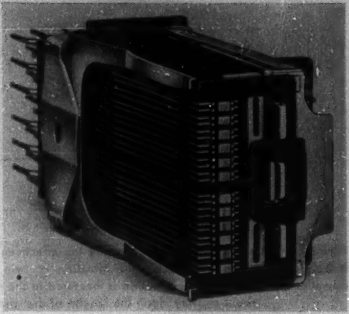


Figure 30 (right). AJ relay with 12 transfer contacts

cuit to limit the current and control the arcing period.

That high weld strengths are obtained by this process is indicated in Figure 15 which shows typical distributions of weld strength for both the percussive-welded contacts and the spot-welded twin contacts. The plots show the percent of contacts with weld strengths equal to or less than any prescribed value within the range of the chart. As shown, the percussive welds are generally stronger than the spot welds which is due, in part, to larger welded areas.

Although percussive welding is more suitable for the stationary contacts welded in the factory, it is planned that occasional replacement of both stationary and twin contacts will be made in the field by spot welding. This will be done with the Bell System field-welding equipment[9] provided with suitable electrodes. In this case, however, more expensive all-palladium stationary contacts of special shape would be used to facilitate the spot welding and individual hand adjustment for final position of the contacts will be necessary.

Standardized Assembly of Coded Parts

Since assembly was one of the most promising fields for reducing costs in a new relay design, special effort was made to reduce the assembly cost of the *AF* relay. Some of the major design features which contribute to low-cost assembly are:

1. The continuous molding and fabricating processes for the wire spring subassemblies, which avoid all individual handling of wires and contacts.
2. Clamping the relay pile-ups by means of a simple spring clamp instead of the more conventional method using screws.
3. A single, easily mounted, operating card.

Less obvious, but equally important, is the basic philosophy whereby a large variety of different relay codes is obtained by assembling parts which for assembly purposes are essentially identical for each code. As previously described, the spring combination for each relay is controlled by selection of the proper code card, twin wire assemblies with wires in the proper positions for that combination, and a stationary wire assembly with the right kind of contacts welded to the proper wires. At the present time 6 different card varieties, 50 twin wire assemblies, and 75 stationary wire assemblies have been standardized. The twin wire assemblies are provided with any number from 1 to 12 pairs of wires in various positions while the stationary wire assemblies have from 1 to 12 contacts in matching positions, with the added variable that each contact may have precious metal on either or both sides as needed. With these it is possible to obtain more than 300 different contact combinations, although only about 100 of these are needed now. Yet, with a few exceptions, each relay code is assembled from the same number of parts put together in the same manner. By using additional varieties of cards and wire spring assemblies the total number of contact combinations which are possible with the basic design is

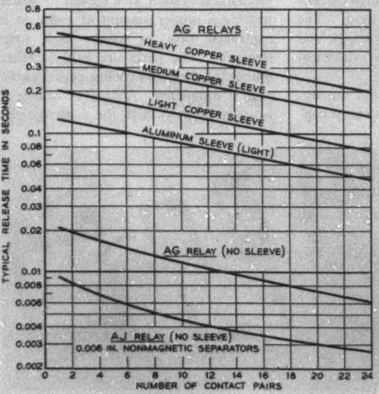


Figure 29 (left). Typical release times of AG and AJ relays

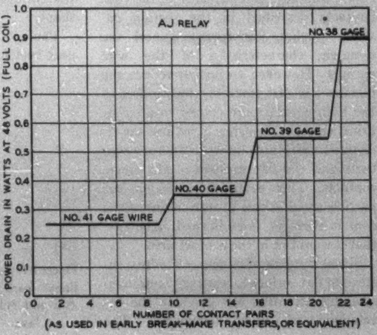


Figure 31 (right). Power used for AJ relays with coils designed for least power

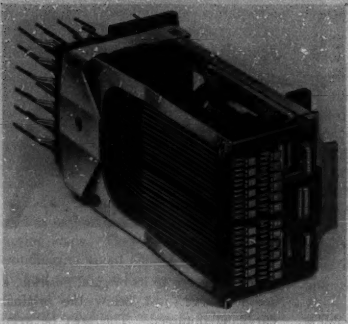


Figure 32 (left). AJ relay with 24 make contacts

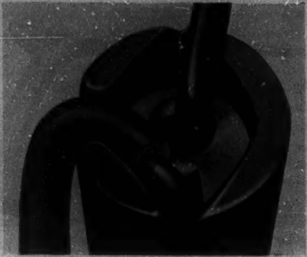


Figure 34 (right). Muzzle of wiring tool showing terminal and connecting wire in position ready for wrapping

many times larger than the 300 just indicated.

Other examples of coded parts which are assembled in a standardized manner are the coils, core plates, and restoring springs. Although coils vary greatly as to turns, resistance, and the like, all are assembled to the cores by the same procedure, using identical spoolheads. The three values of armature travel are controlled by selection of core plates with the proper size openings, but all core plates are assembled alike. Similarly, the restoring springs are provided in seven varieties including six different thicknesses and seven predeflections to give the desired restoring force, but these variations do not affect the assembly operations.

Standardized assembly of coded parts is of value not only in reducing the cost of hand assembly operations but also in providing a more uniform product and as a principle which may make machine assembly practicable.

Relay Adjustment

Since adjustment expense accounts for a considerable part of the manufacturing costs of older type relays, special efforts were made in the design of the *AF* relay to reduce the need for adjustment. As a result several types of adjustment used with other relays have been eliminated completely and the remaining adjustments have been simplified. All individual contact adjustment has been eliminated and only two types of factory adjustments are made with the *AF* relay. These include adjustment of the restoring spring to control armature back tension and a gang adjustment of the stationary contacts to control the points in the armature travel at which the contacts operate. Even these adjustments are needed for only a fraction of the relays, as close control of the tolerances in manufacture often causes the back tension and contact operate points to fall within acceptable limits as the relays are assembled.

The gang adjustment of the stationary contacts is made by bending the arms of the core plate, thereby changing the position of the front molded section of the stationary wire assembly which rests on the ends of the arms. Each arm may be bent by means of a screwdriver inserted in the slot as shown in Figure 16. Rotation of the screwdriver in a counterclockwise direction will cause the upper end of the core plate arm to move to the left, carrying with it the upper end of the stationary wire assembly, including the stationary contacts. This reduces the gap between these stationary contacts and the make twin contacts, thereby causing these contacts to operate earlier in the armature travel. Since the break twin contacts rest against the stationary contacts, these also are moved to the left, reducing the space between the break twin wires and the actuating surface of the card. Thus, bending the core plate arms to the left causes both make and break contacts to operate earlier in the armature travel, while bending the arms to the right causes these contacts to operate later. By bending both arms in the same direction, the operate points of all contacts may be shifted in the same direction. On the other hand, separate arms permit adjustment of the upper relay contacts independently of the lower contacts, thereby increasing the latitude of adjustment.

The parts of the relay are dimensioned so that no adjustment of the core plate arms is required, except to compensate for variations in manufacture of the relay parts. Hence relays assembled from parts made with sufficient accuracy generally do not require adjustment.

The restoring springs may be adjusted for the proper armature back tension by

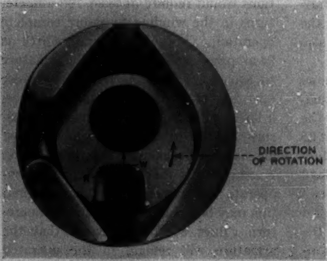


Figure 33 (left). Muzzle of wiring tool for making wrapped connections

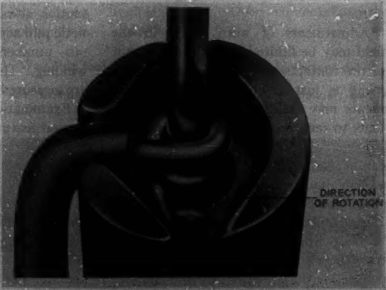


Figure 35 (right). Muzzle of wiring tool showing two turns wrapped by rotation of the inner cylinder

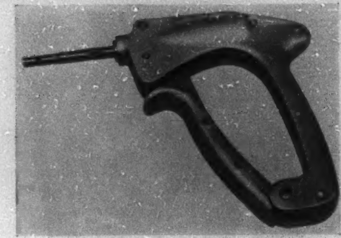


Figure 36 (left). Hand-operated wrapping tool for installation and repair service

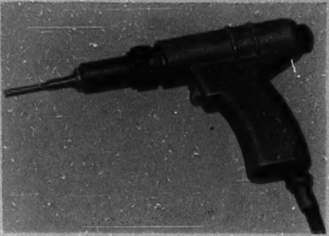


Figure 37 (right). Wrapping tool operated by air pressure for factory use

the use of a simple spring bending too applied to the side arms. However, springs are provided with various predeflections and thicknesses to correspond with various numbers of make twin contacts which must be held open in the unoperated position. Again, no adjustment for back tension is necessary except to compensate for variations in manufacture as the restoring spring tension is normally just sufficient to overcome the tension of the make twin wires and hold the armature against the backstop within acceptable force limits. Close control of the tension bends in the wires and restoring springs reduces the frequency with which adjustments are needed and a large portion of the relays do not require this adjustment.

Types of factory adjustment which are common on other relays but which have been eliminated entirely on the *AF* relay include adjustments for contact force, individual adjustment of contacts for contact operate point, and adjustment for armature travel. Contact force is controlled by means of the large predeflections of the twin wires as mentioned previously. Individual contact adjustment is eliminated by close control of tolerances combined with the single card method of actuation, and by the simpler gang adjustment used when necessary. Adjustment for armature travel is eliminated by the use of close tolerances on the controlling dimensions of the various parts.

Adjustments of worn relays in the field may be limited to gang adjustment of the contacts and back tension adjustment as just described. Other adjustments may include burnishing the contacts to remove surface irregularities, replacement of contacts and individual contact adjustment as mentioned previously, and replacement of the card if it should become badly worn or damaged. If card replacement is necessary, this may be done without dismounting the relay from the mounting plate and without disconnecting the associated wiring.

Relay Performance

As previously stated, the broad objective in the design of the *AF* relay has been to reduce the annual charges for the use of this relay in the telephone system. Part of this reduction comes from lower manufacturing costs; the remainder comes from savings associated with the improved performance characteristics, such as long life with relatively low maintenance expense, reduced power consumption, and increased speed which reduces the number of units of certain types of equipment, such as markers, needed for telephone central offices. A brief description of some of the principal characteristics of the new relay follows.

Load and Pull Characteristics

Figure 17 shows typical load and pull curves for a wire spring relay with 12 early break-make transfer contacts. The abscissa shows the motion of the armature as it travels from the unoperated position to the operated position, and back again. This is measured at the centerline of the card and hence is also the card motion. In the unoperated position the armature rests against a backstop, which is part of the core plate. In the operated position it rests against 0.006-inch-thick nonmagnetic separators which prevent the armature from touching the core. The ordinate shows the spring load, which opposes the armature motion toward the core, and the magnetic pull acting on the armature for various numbers of ampere-turns in the winding. These pull and load curves also are measured at the card.

Examination of the load curves shows several features of the relay. The armature back tension, or force, holding the armature against the backstop is about 65 grams in this case. As the armature moves toward the core, the spring load increases along the upper of the two nearly parallel load curves until it reaches a final value of about 440 grams in the operated position. As the armature is allowed to return to its original position, a second curve, just below the original curve, is obtained. The area between these two curves is a measure of the energy loss due to mechanical hysteresis, or friction, in the relay. As can be seen from the curves, the friction in the new relay is very low and is a small fraction of the spring load at all values of armature travel.

The shape of the load curves is characteristic of *AF* relays with intermediate travel (0.044 inch). The load increases rapidly in two regions, corresponding to the intervals in which the early and late contacts operate. The rapid increases are caused by the armature and card picking up the additional load of the twin wire springs. Each of the 48 twin wires is picked up almost abruptly at various points and the summation of these additions to the load gives the irregular appearance shown.

The pull curves of Figure 17 are for essentially static conditions since the armature was restrained to move slowly through its travel while the curves were automatically recorded. These curves are of interest because they show the ampere-turns necessary to assure operation of the relay and also values which will assure the armature will not leave the backstop. For example, the "critical load point," or point on the load curve which requires the greatest number of ampere-turns, is seen to occur at 0.025 inch travel and 250 grams, which under static conditions would require at least 160 ampere-turns in the winding to assure complete operation. On the other hand, as little as 94 ampere-turns could cause the armature to leave the backstop and might cause operation of one or two contacts. Hence, a lower value must be maintained to assure that the armature will remain at rest against the backstop. This information is important for relays having nonoperate requirements. Similar information may be obtained for limiting ampere-turn values which will assure that the armature will remain in the operated position (hold requirements) and, again, which will assure complete release to the backstop position (release requirements).

Speed of Operation and Release

Typical operate and release times of the *AF* relay are shown in Figures 18 and 19. Figure 18 shows the operate times for "speed relays" in which the speed is limited primarily by the time needed to accelerate the mass of the moving system, and is not affected appreciably by the spring loads. These are relays with coil windings of about 1,000 ohms resistance, or less, corresponding to power inputs of 2.3 watts, or greater, when connected to a 48-volt supply. For each value of resistance, the operate time shown is obtained with windings having the optimum number of turns for shortest operate time. This time is plotted as a function of power input for various armature travels. Short travel relays have typical operate times varying from about 2.5 to 7 milliseconds as the power is reduced from 23 to 2.3 watts, or as the resistance is increased from 100 to 1,000 ohms, using the appropriate number of turns in each case. Intermediate and long travel relays have longer times. For comparison, a short travel *U* relay is also shown. This relay requires about twice the time to operate as the corresponding *AF* relay. The improvement with the *AF* relay is due primarily to the lighter mass of the moving system and slightly shorter travel due to better control of tolerances. Better control is inherent with the single card system for contact actuation and is accomplished without individual adjustment of contact or backstop position, both of which are hand-adjusted on *U* relays.

Typical release times for the *AF* and *U* relays are shown in Figure 19, with time plotted as a function of the number of contact pairs for relays equipped with standard 0.006-inch-thick nonmagnetic separators. In this case the improvement is greater than two to one, due principally to the lighter moving parts of the *AF* relay and lower eddy current effects of the rectangular silicon iron core.

Power Requirements

Figure 20 shows the nominal power required to assure operation, with some margin, of relays with windings designed for minimum power consumption. Included is an allowance for adverse variations in magnetic structure, winding, and loads. Since least power will be used by the largest coil wound with the finest wire consistent with meeting the ampere-turn requirements for the various contact loads, the curves are discontinuous and have steps as the wire sizes are shifted from one size to the next to meet the ampere-turns needed for increasing loads. Again, the corresponding *U* relay power is shown for comparison. For corresponding numbers of contact pairs, the *AF* relay requires about half the power of the *U* relay, except with fewer contact pairs where the power in each case depends upon the use of number 41 gauge wire. This comparison applies only when coils of optimum design for minimum power are used on both relays. In practice, the coils are selected for the best economic balance between power consumption, cost of the coils, and value of speed of operation. Coils designed for minimum power are relatively expensive because they contain as many turns of fine wire as the available space permits, and their use is economical only on relays which are operated an appreciable portion of the time and where speed is relatively unimportant. The reduced power required for the *AF* relay is due principally to an improved magnetic structure, shorter travels for similar contact combinations, and lower contact forces.

Contact Performance

The principal characteristics which must be considered in evaluating contact performance include chatter, erosion or wear, susceptibility to open contacts and locking, and changes in these characteristics with wear of the relays. In general, all these features are improved on the *AF* relay compared with the *U* type.

Figure 21 shows typical chatter on closure of make and break contacts on *U* and *AF* relays built for moderate and fast operation. The degree of improvement of the *AF* relay is striking. The reduction in chatter has direct circuit advantages in reducing the possibility of false operation of associated high-speed equipment and also indirect advantages in prolonging the life of the contacts. The improvement is due largely to the type of card operation of the completely independent, low-mass twin wires and also to the low mass of the moving system which excites less vibration of the relay structure as a whole. The placement of the card close to the contacts allows the full contact force to be developed within a very short time, and the low mass of the twin wires stores little kinetic energy to cause reopening due to wire vibration.

A particularly troublesome type of chatter occasionally experienced is caused by rebound of the armature after striking the backstop. This chatter is objectionable because of its long duration which is of the order of a millisecond and may occur several milliseconds after the initial opening or closure of the contacts. This increases the possibility of false operation of associated circuits. Accordingly, a fundamental study was made of the means for reducing armature rebound, as previously mentioned. As a result, changes were made in the suspension of the armature and in the position of the backstop which substantially eliminated chatter due to armature rebound.

Electrical erosion of the contacts is reduced on the *AF* relay because of less chatter and because of the lower energy levels controlled by the contacts, where these are used to operate other *AF* relays. This improved performance not only reduces maintenance but permits the use of less expensive, smaller sized contacts.

Contact locking is substantially eliminated on the *AF* relay because of the card operation, where the static and dynamic forces associated with the card and armature are available to break loose any incipient lock. Open contacts are reduced by

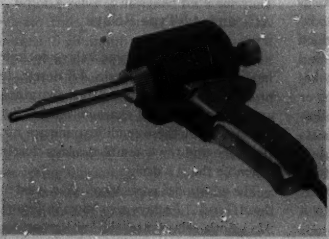
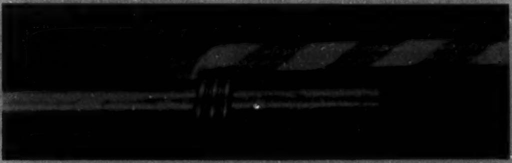


Figure 38 (left). Wrapping tool operated by electric motor for factory use

Figure 39 (below). A wrapped connection on a wire spring terminal, before soldering



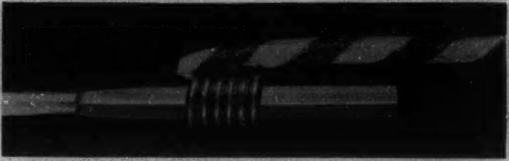


Figure 40. A solderless wrapped connection on a wire spring terminal

1. protecting the contacts from dust with a small cover,
2. rounding the twin contact surfaces to to reduce the effective areas on which particles must lodge to cause opens and to increase the pressure on the areas,
3. the use of palladium contacts,
4. the dynamic characteristics of the wire springs, and
5. the use of twin contacts on completely independent twin wires.

The complete separation of the twin wires is an important feature in reducing open contacts. As shown in Figure 22, the flat punched springs of the *U* relay carry twin contacts but these are mounted on tips which are separated by a relatively short punched cutout. This limited separation did not achieve the full advantage of twin contacts as a sufficiently large particle of dust under one contact could cause both contacts to be held open. A subsequent design, known as the *UB* relay,[9] used a longer cutout, resulting in greater independence of the twin contacts with a significant reduction in contact opens. The *AF* wire spring relay achieves complete separation by the use of separate twin wires and a significant part of the improvement with respect to contact opens is due to this feature.

The improvements in contact chatter and open contacts become even more evident during the life of the relay. As shown in Figure 23, the contact forces on *U* relays diminish rapidly with wear of the contacts resulting in increased chatter, more frequent opens, and the need for earlier readjustment. Card operation of wire springs with large predeflections, however, assures substantially constant forces, thereby maintaining the initial chatter-free performance and fewer open contacts.

LIFE

Tests of relays with contacts protected electrically with resistance-capacitance networks indicate that the standard *AF* relays with winding resistances of 700 ohms or greater will have a life in the order of 250,000,000 to 500,000,000 operations before readjustment becomes necessary. With readjustment, of course, these figures can be increased several times.

Where longer life is essential, special features are used to increase the life. With these features a life, before readjustment, of a billion operations is expected for some relays and, with readjustment, all relays equipped with the special features for long life should be capable of a billion operations.

The special features for long life include vibration dampers attached to the twin wires and the stationary wire assembly as shown in Figure 24, wear-resistant nickel and chromium plate on the armature, core, and core plate, and a long-wearing alloy for the nonmagnetic separators welded to the armature. The special features consist largely of variations in finish and material which do not greatly affect the manufacturing processes. The only added parts are the damping members. These are molded from soft but stable polyisobutylene with grooves to receive the twin wires. One damper is attached to each side of the shelf provided on the stationary wire assembly. The twin wires pass through the grooves and are cemented in place. As shown in Figure 25, these dampers reduce the vibration of the twin wires between the card and the molded section at the rear, thereby reducing the slide between the wires and the card.

Early designs of relays indicated that wear between the twin wires and the card was excessive and that changes in materials would not produce the improvement needed for very long life, particularly with high-speed relays. A fundamental study[10] of the conditions which cause wear was made and it was found that reduction of the sliding motion between the wires and card to 0.001 inch or less was necessary to substantially eliminate such wear. The *AF* relay meets this requirement. The necessity for such a requirement will be better understood when it is noted that, for one billion operations, the total slide corresponds to a distance of about 32 miles.

STABILITY

The *AF* relay is a distinct improvement in stability compared with earlier designs when subjected to shock or temperature and humidity changes. Under severe and repeated variations in temperature and humidity, the largest changes in contact separation are not more than 0.002 to 0.003 inch. The improved stability is expected to permit final adjustment and inspection of the relay at the time it is assembled without need for readjustment after it is wired into equipment and installed into service after shipment.

MAGNETIC INTERFERENCE

In the past it often has been necessary to maintain large spacing between relays where critical values of current to operate or release the relays must be maintained. In some cases special iron shields were used for further magnetic isolation. Without these precautions, the leakage flux from adjacent relays entered the magnetic circuit of the critical relays and the operate or release currents varied according to whether the adjacent relays were energized.

Magnetic interference between *AF* relays is substantially eliminated as shown in Figure 26. This is largely because of the low reluctance of the magnetic circuit resulting from the 1-piece core and the large pole-face areas between the core and armature. As shown in the figure, the measurements were made by surrounding a control relay with eight adjacent closely spaced interfering relays. The ampere-turns to just operate the control relay were varied by changing the mechanical load on the relay, and for each value the change caused by simultaneously energizing the adjacent relays was observed. The improvement of *AF* relays with respect to the *U* type is seen to be of the order of ten times for most of the range, with the effects of the adjacent relays being well under 10 per cent up to 300 ampere-turns. This is small enough so that no shields or precautions with respect to spacing are required.

AG- and *AJ*-Type Relays

The *AG*- and *AJ*-type relays include modifications of the basic *AF* design to provide slow release, sensitive, marginal, and other additional characteristics For the most part these modifications are not extensive and the assembled relays closely resemble the *AF* design.

The additional parts most often used in the *AG* and *AJ* relays are shown in Figure 27. Both relays use thicker armatures with longer side legs than the *AF* relay, and the armature of the *AG* relay has a spherical embossing instead of nonmagnetic separators. This reduces the magnetic circuit reluctance of the *AG* relay when it is in the operated position. In addition, for longer release times, a metal sleeve is assembled over the middle leg of the core, inside the coil. Induced eddy currents in this sleeve oppose rapid changes of flux through the core.

The use of the heavy, embossed armature and a sleeve are sufficient to make the relay slow to release.[11] When the current in the winding of such a relay is interrupted, the flux decays slowly due to the circulating currents in the sleeve. Also, the low magnetic reluctance increases the time for the flux decay by permitting relatively high flux values to be maintained by smaller circulating currents. However, to achieve better control of the release times and to maintain stable adjustment during the life of the relays, the following additional features are used:

1. The cores are annealed in a hydrogen atmosphere, chiefly to stabilize the coercive force of the iron.
2. The core and armature have a wear-resisting chromium plate finish to maintain the nonmagnetic gap between the embossed surface of the armature and the core.
3. The use of a spherical embossing reduces variations in reluctance caused by small angular misalignments between the armature and core.
4. Four sleeves are available including light, medium, and heavy copper sleeves and a light aluminum sleeve. These sleeves provide various ranges of release time.
5. A buffer spring is provided on the relay to control the operated load and therefore the release time. As shown in Figure 28, the buffer spring normally is tensioned against the end of the middle core leg. As the relay operates, however, the card strikes the adjustable tab in front of the middle leg and lifts the spring away from the core so that the spring tension is added to the operated load of the relay. The spring may be adjusted for any desired tension, within limits, and the tab can be adjusted so that the load may be picked up at any desired point in the armature travel.

When a relay is designed for a specified release time, the spread between maximum and minimum times obtained with a particular sleeve is usually greater than desired. Accordingly a sleeve is selected which under normal conditions would produce a somewhat longer time than the specified value. This time then is reduced, as needed, by increasing the buffer spring tension.

Figure 29 shows typical release times plotted as a function of the contact load for *AG* relays with and without sleeves. Since this figure illustrates release times that are characteristic of the various sleeves, no buffer spring tension is assumed. As would be expected, the heavier sleeves produce the longer release times, which are also greater for relays with fewer contacts. Even the light aluminum sleeve produces several times longer release times than no sleeve at all. For comparison, release times also are shown for the *AJ* relay, which has a magnetic structure similar to the *AG* but with nonmagnetic separators in place of the spherical embossing on the armature. The difference between the *AJ* release times and those for the *AG* relay with no sleeve shows the time advantage obtained with the domed armature.

The *AJ* relay, with its long and relatively heavy armature, is suited for the more critical marginal applications and is capable of operating heavier contact loads than the *AF* relay. All relays with more than 18 contact pairs are provided only on the *AJ* structure. For example, Figure 30 illustrates an *AJ* relay with a full complement of 12 transfers (24 contact pairs).

A measure of the power requirements for the *AJ* relay is given in Figure 31. This shows the power required to assure operation with various numbers of contact pairs for coils designed to consume minimum power at 48 volts. The chart includes allowances for variations in load, magnetic structure, and coil, with some margin for changes in these characteristics. Comparison with Figure 20 shows the power requirements to be slightly lower than for the *AF* relay except for small numbers of contacts where limitation of wire sizes of the coils is controlling. Under limiting conditions the *AJ* relay will operate on as little as 0.025 watt.

Other features which may be used to extend the use of the *AJ* relay for special marginal applications include armatures with various thicknesses of nonmagnetic separators, wire spring assemblies with reduced contact forces, core laminations (strips of iron placed inside the coil, against the middle leg of the core to increase the effective cross-section), and the buffer spring which may be used to control the operated load of the relay and therefore the hold and release currents.

A special variety of *AJ* relay is provided with 24 pairs of make contacts, as shown in Figure 32. This relay uses four layers of wire springs and a number of other special parts. As a result, it is often possible to use one 24-make contact *AJ* relay rather than two relays with fewer contacts on each.

Wiring the Relays

Connecting wires to the wire spring relay terminals presented a problem which was solved not only for the new relay but by methods which have become important and useful for other apparatus also. The solution came from the invention of a tool, first proposed by H. A. Miloche,[12] for quickly and easily wrapping the connecting wire to the straight end of the wire spring relay terminal.

Early suggestions for making connections to the wire springs included various hooks or bends to simulate the common flat punched terminal having either a hole or notch to facilitate attaching the wires. All of these added some expense to the manufacture of the relay. The added costs were due to two factors:

1. Forming the wire spring ends required additional operations.
2. The greater flexibility of the wire spring terminals caused some difficulty for the operator so that some increase resulted in the time required to make a connection.

The wrapping tool first was visualized as a simple trigger-operated hand tool, later as an air or electrically operated tool, and still later as a combination tool to do additional operations including cutting and skinning the connecting wire.

Although the wrapping tool was developed to solve a problem which arose in the development of the wire spring relay, it first was applied in commercial practice by the Western Electric Company for wiring to the flat spring relay terminals. Wrapped connections are used extensively now with these terminals, resulting in an improved product at a lower cost. In making wrapped connections to either flat or round terminals, it was expected that tinned terminals and wire would be used and soldered together to give a stable, low-resistance junction. More recently, however, it has been possible to show that soldering is not required if certain definite dimensional conditions are satisfied by the terminal and the wrapping tool. Solderless wrapped connections will be described in detail in a separate paper.

BASIC PRINCIPLES OF WRAPPING TOOLS

Figure 33 shows a drawing of the arrangement and action of a wire wrapping tool. There are a number of dimensions of the tool and of the terminal which have engineering importance. For the purposes of this paper, it will suffice to note that the radius *R* and the wall thickness *W* are important in producing the best wrapped connection.

As shown in Figures 34 and 35, the inner member of the tool rotates around the terminal which is inserted in hole *A* of the rotating member. The connecting wire is inserted at *B* and is anchored by a slight force against the outer stationary member of the tool. As the inner member rotates, the connecting wire is stretched and formed around the terminal until all of the wire length is used. It should be noted particularly that all of the wire is used, making it unnecessary to clip a wire end as in other wiring methods. This is an important detail in avoiding wire clippings which sometimes cause unwanted cross connections in wired equipment units. The tool tip described can be operated by a hand trigger or by motor. Figure 36 shows a hand-powered tool primarily for installation and repair service. Figure 37 shows a production-type tool driven by air pressure developed by the Western Electric Company at the Hawthorne Works. Figure 38 shows a tool driven by an electric motor used by the Kearny Works of the Western Electric Company. Figure 39 shows a wrapped connection on a wire spring relay prior to soldering.

Figure 40 shows another wrapped connection of the solderless type on a similar terminal. Studies indicate that solderless wrapped connections can be used with a wide variety of materials, including aluminum.

It is interesting to note that a troublesome problem in wiring to the new wire spring relay was solved by the development of a new method which itself has become an important development with broad applications. The wrapped connection with or without subsequent soldering has resulted in better, more uniform, and less costly connections made in less time than those made by previous methods.

Conclusions

A description has been given of a new type of wire spring general-purpose relay for telephone switching systems. Although accurate manufacturing costs will not be available for some time, the new relay is expected to be substantially lower in cost. It provides major improvements in contact performance, reduced power, faster operation, and longer life. The new relay also covers a wider field of application than any previous general-purpose relay in such characteristics as speed, slow release, marginal operation, number of contacts, and so forth.

Important economic advantages include lower manufacturing cost of the relay itself and a reduction in switching systems costs resulting from less equipment and reduced power.

The new relay has shown considerable improvement in mechanical life and in contact performance. The life of the relay is of the order of one billion operations. These improvements can be expected to reduce the cost of maintenance of switching systems appreciably.

References

1. Relays in the Bell System, **S. P. Shackleton, H. W. Purcell.** *Bell System Technical Journal* (New York, N. Y.), January 1924, page 1.
2. The U-Type Relay, **H. N. Wagar.** *Bell Laboratories Record* (New York, N. Y.), May 1938, page 300.
3. The UB Relay, **H. M. Knapp.** *Bell Laboratories Record* (New York, N. Y.), October 1949, page 355.
4. The Y-Type Relay, **F. A. Zupa.** *Bell Laboratories Record* (New York, N. Y.), May 1938, page 310.
5. Improved U, UA and Y Type Relays, **W. C. Slauson.** *Bell Laboratories Record* (New York, N. Y.), October 1951, page 466.
6. Internal Temperatures of Relay Windings, **R. L. Peek, Jr.** *Bell System Technical Journal* (New York, N. Y.), January 1951, page 141.
7. Cellulose Acetate Filled Coils, **C. Schneider.** *Bell Laboratories Record* (New York, N. Y.), November 1951, page 514.
8. Relay Armature Rebound Analysis, **R. E. Sumner.** *Bell System Technical Journal* (New York, N. Y.), January 1952, page 172.
9. Relay Contact Welder, **W. T. Pritchard.** *Bell Laboratories Record* (New York, N. Y.), April 1944, page 374.
10. New Techniques for Measuring Forces and Wear in Telephone Switching Apparatus, **W. P. Mason, S. D. White.** *Bell System Technical Journal* (New York, N. Y.), May 1952, page 469.
11. Slow Acting Relays, **H. N. Wagar.** *Bell Laboratories Record* (New York, N. Y.), April 1948, page 161.
12. Mechanically Wrapped Connections, **H. A. Miloche.** *Bell Laboratories Record* (New York, N. Y.), July 1951, page 307.

No Discussion

Type-O Carrier Telephone

J. A. COY
MEMBER AIEE

E. K. VAN TASSEL
NONMEMBER AIEE

TYPE-*O* carrier, the most recent addition to the carrier series, is designed to supply additional telephone circuits to the many communities served by open-wire facilities. This new system supplements, rather than supersedes, the older open-wire systems such as types *C*, *H*, and *J*. Type-C and *J* carrier systems were designed primarily for long circuits, 2,000 miles or more, and consequently cost more than buying new wire for short distances. This leaves an important field of use for an economical short-haul carrier system, particularly below 150 miles.

Type-*O* carrier is a family of four 4-channel systems providing 16 two-way talking channels on a single open-wire pair. These systems, individually known as *OA*, *OB*, *OC*, and *OD*, are suitable for use over distances up to 150 miles and will prove in for distances as short as 15 or 20 miles, the exact distance depending on local conditions such as terrain, population density, and the amount of unfilled space on existing pole lines. Each system occupies a band of frequencies 36 kc wide as shown in Figure 1. The *OA* system is equivalent to a voice circuit plus a type-*C* system and has a top frequency of 36 kc. The *OB* system, located in the band 40 to 76 kc can be operated above a type-*C* system without any change in the line transpositions. The *OC* and *OD* systems are high enough in frequency so that additional line transpositions are usually required.

This new type-*O* system has been made economical for short-haul use without sacrificing high standards of circuit reliability and quality, by designing it specifically for short distances, by employing built-in compandors and built-in signaling, by the use of frogging repeaters so that only flat gain regulation is needed, by new apparatus components not previously available, by assembly-line manufacture and the use of die casting for all framework parts, and by packaging for minimum application engineering and installation effort.

Paper 52-307, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted March 24, 1952; made available for printing September 4, 1952.

J. A. Coy and E. K. Van Tassel are both with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

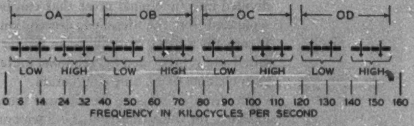

OA
OB
OC
OD
LOW
HIGH
LOW
HIGH
LOW
HIGH
LOW
HIGH
0 8 14 24 32 40 50 60 70 80 90 100 110 120 130 140 150 160
FREQUENCY IN KILOCYCLES PER SECOND

Figure 1 (above). O carrier telephone frequency allocation

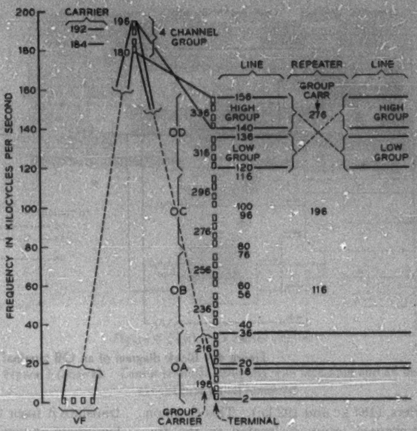

200
180
160
140
120
100
80
60
40
20
0
FREQUENCY IN KILOCYCLES PER SECOND
CARRIER
192
184
196
180
4 CHANNEL GROUP
LINE
REPEATER
LINE
GROUP CARR
276
156
HIGH GROUP
140
136
LOW GROUP
120
116
HIGH GROUP
LOW GROUP
336
316
296
276
256
236
216
196
OD
OC
OB
OA
100
96
80
76
60
56
40
36
20
16
2
196
116
VF
GROUP CARRIER
TERMINAL

Figure 2 (right). Type-O carrier modulation plan

Type-*O* carrier systems fulfill the same purpose on open-wire lines as type-*N* carrier systems[1] do on cable routes. Many features, such as the built-in compandor and the built-in signaling which provides for both dialing and supervision, are common to both types of systems. These features have been described in some detail and consequently need only be mentioned here.

The outstanding differences between type-*O* and type-*N* systems are: The transmission of two channels as the sidebands of one carrier instead of double side-band transmission; transmission in both directions on one pair instead of separately on two pairs; a range of regulation four times that required in type *N*.

The combination of two voice circuits on one carrier is referred to as twin-channel operation. This type of operation is an efficient use of the frequency space, requires the generation of only one carrier frequency per two channels, and permits the gain regulation of twin channels by a common circuit. The discrimination requirements in the receiving band filter are made more lenient by this twin-channel operation since the carriers are 8 kc apart and the "unwanted" energy that gets through the band filter from some other twin will demodulate as frequencies above 4 kc. These can be removed in a simple voice frequency roof filter rather than by more discrimination in the channel band filter.

Combining the two directions of transmission on one pair introduces the need for more filters to separate and combine groups than are used in type *N*. An outstanding feature of type *O* is the way in which filters are used so as to reduce the number of codes and increase the flexibility of the system. Filters generally are connected through plugs and jacks so that they can be rotated 180 degrees or exchanged for other types, to produce different circuit arrangements.

In a type-*N* system the primary source of change in cable attenuation is the change in resistance due to variations in the temperature of the copper wires. In an open-wire line this factor is small compared with the extra loss inserted by leakage across wet surfaces during rain. Also the presence of sleet on the wires increases the loss many times. The regulating range therefore must be three or four times greater in an open-wire system than in a cable carrier system.

Each of the type-*O* systems starts with a 4-channel bank in which four separate voice circuits are modulated onto two car-

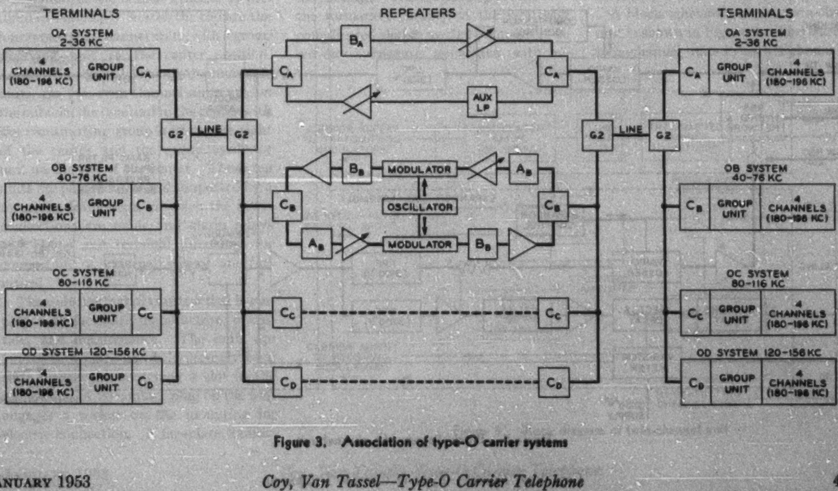

TERMINALS
OA SYSTEM 2-36 KC
4 CHANNELS (180-196 KC)
GROUP UNIT
C_A
OB SYSTEM 40-76 KC
C_B
OC SYSTEM 80-116 KC
C_C
OD SYSTEM 120-156 KC
C_D
G2
LINE
G2
REPEATERS
B_A
AUX LP
B_B
MODULATOR
A_B
OSCILLATOR
MODULATOR
TERMINALS

Figure 3. Association of type-O carrier systems

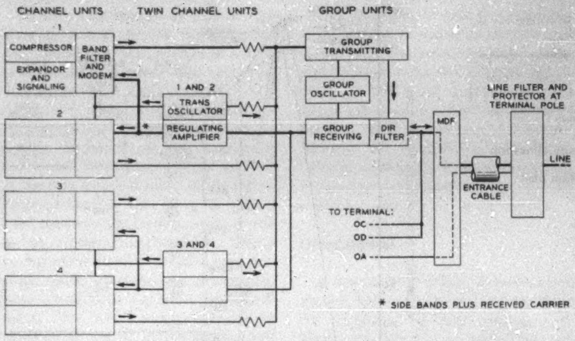

CHANNEL UNITS
TWIN CHANNEL UNITS
GROUP UNITS
1
COMPRESSOR
EXPANDOR AND SIGNALING
BAND FILTER AND MODEM
1 AND 2
TRANS OSCILLATOR
REGULATING AMPLIFIER
2
3
3 AND 4
4
GROUP TRANSMITTING
GROUP OSCILLATOR
GROUP RECEIVING
DIR FILTER
MDF
LINE FILTER AND PROTECTOR AT TERMINAL POLE
LINE
ENTRANCE CABLE
TO TERMINAL:
OC
OD
OA
* SIDE BANDS PLUS RECEIVED CARRIER

Figure 4. Block diagram of an OB terminal

riers, (184 kc and 192 kc). This common 4-channel group is located in the frequency band 180 to 196 kc. This band is then group modulated into any one of the desired 16-kc bands, as shown in Figure 2, for transmission over the line.

At repeater points the *OA* signals (2 to 18 kc and 20 to 36 kc) are amplified and transmitted without any change in frequency. In this way type -*OA* signals experience the same gains and losses as a type-*C* system on the same pole line, the ideal condition for minimum crosstalk between the two systems. The *OB*, *OC*, and *OD* signals at repeater points are all "frogged," that is each 16-kc band is group modulated so as to invert it and translate it from the high group into the low group or vice versa, before it is amplified and transmitted over the line.

At the receiving terminal all of the 4-channel groups are group modulated back to the 180- to 196-kc range, separated by the receiving channel band filters, and demodulated to voice frequencies.

Terminal

With that brief outline of the modulation steps used in the type-*O* system a more detailed description can be made starting with a terminal and proceeding to the repeater. Figure 3 is a block schematic of the four type-*O* systems, divided into functional areas that can be separated for the detailed description. Figure 4 is a block diagram of an *OB* terminal and it is divided into channel units, twin-channel units, and group units.

Provision is made in the channel unit, Figure 5, for the connection of a 2-wire or a 4-wire voice frequency circuit. The normal input from the connecting circuit measured at the 4-wire point is at a −16-decibel level. These voice-frequency signals are first compressed by a two-to-one compressor in which the volume range of the output signal is 1/2 the decibel volume range of the input signal. After passing through a low-pass filter to suppress frequencies above 3,100 cycles the compressed speech is applied to the balanced modulator. The 3,700-cycle supervisory and dialing tone which is turned on and off by the signaling keyer circuit also is introduced at this modulator input. This makes the wanted signal the combination of a voice band 200 to 3,100 cycles plus a 3,700-cycle tone. In the modulator output appear the two side bands; the carrier is very small, having been suppressed by the modulator balance. A band-pass filter further attenuates the carrier and suppresses the unwanted side band. The wanted side band is combined with the three side bands from the other 3 channels and the carriers are reinserted to form the 4-channel group output.

The group transmitting unit takes this 4-channel group and, depending upon the group oscillator supplied, group modulates these channels to the desired line fre-

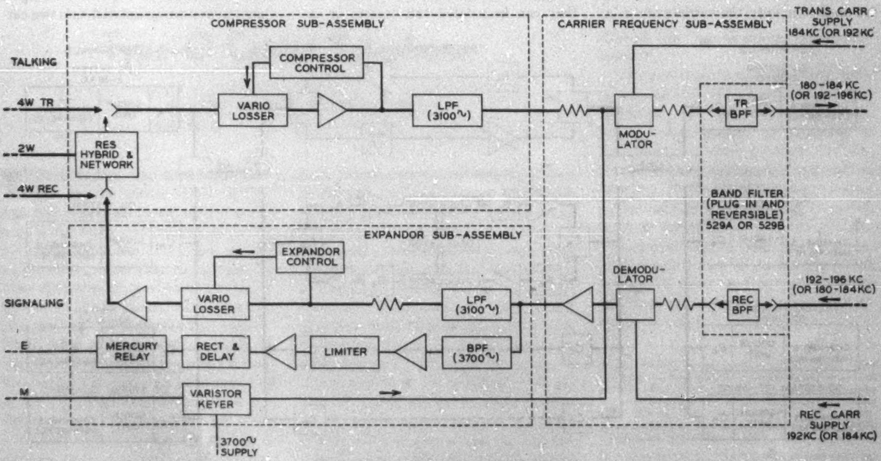
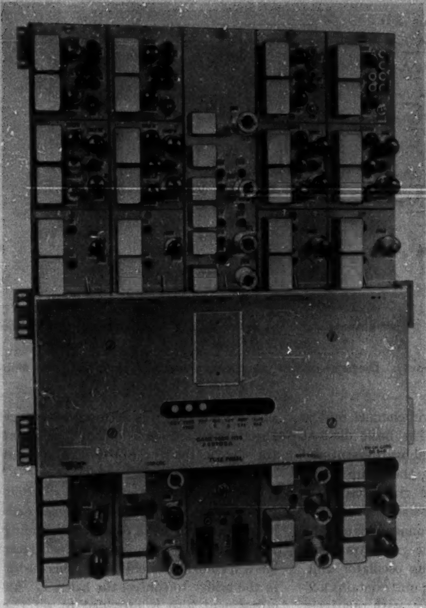

COMPRESSOR SUB-ASSEMBLY
CARRIER FREQUENCY SUB-ASSEMBLY
TRANS CARR SUPPLY 184KC (OR 192KC)
TALKING
4W TR
2W
4W REC
COMPRESSOR CONTROL
VARIO LOSSER
LPF (3100~)
RES HYBRID & NETWORK
MODULATOR
TR BPF
180-184 KC (OR 192-196KC)
BAND FILTER (PLUG IN AND REVERSIBLE) 529A OR 529B
EXPANDOR SUB-ASSEMBLY
EXPANDOR CONTROL
DEMODULATOR
SIGNALING
VARIO LOSSER
LPF (3100~)
REC BPF
192-196 KC (OR 180-184KC)
E
MERCURY RELAY
RECT & DELAY
LIMITER
BPF (3700~)
M
VARISTOR KEYER
3700~ SUPPLY
REC CARR SUPPLY 192KC (OR 184KC)

Figure 5. Block diagram of a channel unit



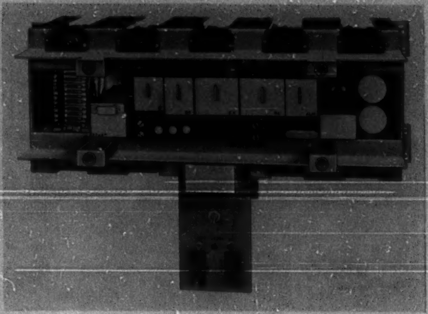


Figure 6 (left). O carrier terminal

Figure 7 (above). Carrier terminal mounting with faceplate and all units except fuse panel removed

quency. The lower side band is selected by the modulator filter and then amplified by a broad-band amplifier. By making this amplifier have adequate gain across the band 40 to 160 kc this transmitting unit can be used in all the *OB*, *OC*, and *OD* systems.

A complete 4-channel terminal is pictured in Figure 6. Across the top are the four removable channel units, with a group receiving unit in the center position. Across the bottom are five removable units, the two twin-channel units are on the outside, the fuse unit in the center with the transmitting group unit on the right of the center and the group oscillator unit on the left of the center. These ten units are plugged into and supported by a mounting framework in which the interconnections are made and alarm relays and lamps and terminal punchings for connection of external wiring are furnished.

The unit method of construction is employed to facilitate manufacture, operation, and maintenance. The units are guided into place and held securely by a slide in the unit matching a slot in the mounting. A 20-contact plug on the unit engages a socket on the mounting for electric connection. A faceplate, removable by means of four fasteners, holds the units in the frame and covers the relays, wiring, and miscellaneous apparatus. Figure 7 shows a terminal mounting with the faceplate and all units except the fuse unit removed. The unit frameworks and the mounting shelves are all aluminum die castings. The die-casting art permits the formation cheaply of the relatively complicated shapes needed to mount present-day miniature apparatus with efficient use of space, accessibility of parts for manufacture and replacement, and a minimum of brackets, screws, and similar miscellaneous items. Some of the larger units have been broken down into subassemblies, three in the channel unit, to obtain sizes convenient for bench work in the shop. These are connected together by multicontact plugs and jacks. Pigtail apparatus is mounted on parallel thermoplastic strips by the same methods as in type-*N* carrier. These assemblies are visible in Figures 9 and 12.

These units and subassemblies are convenient in the field since they can be removed readily when defective, replaced by spares and sent to central points for efficient repair.

A block schematic of the twin-channel unit is shown in Figure 8. There are two twin-channel units in each 4-channel ter-

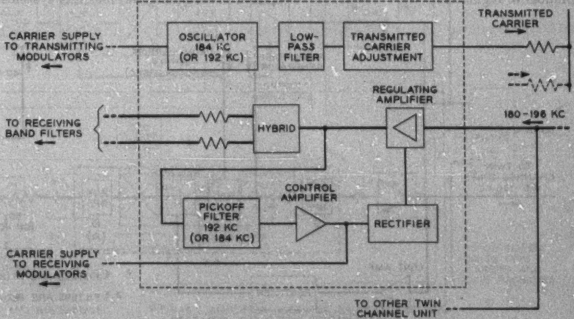
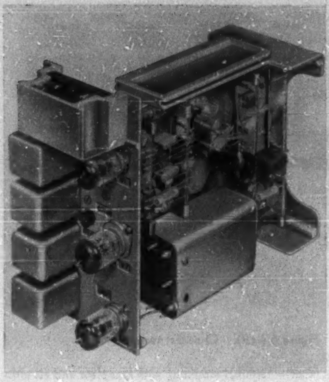

CARRIER SUPPLY TO TRANSMITTING MODULATORS
OSCILLATOR 184 KC (OR 192 KC)
LOW-PASS FILTER
TRANSMITTED CARRIER ADJUSTMENT
TRANSMITTED CARRIER
REGULATING AMPLIFIER
180-198 KC
TO RECEIVING BAND FILTERS
HYBRID
CONTROL AMPLIFIER
PICKOFF FILTER 192 KC (OR 184 KC)
RECTIFIER
CARRIER SUPPLY TO RECEIVING MODULATORS
TO OTHER TWIN CHANNEL UNIT

Figure 8. Block diagram of twin-channel unit



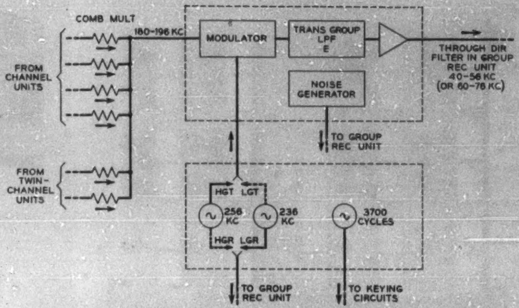

COMB MULT
180-196 KC
FROM CHANNEL UNITS
FROM TWIN-CHANNEL UNITS
MODULATOR
TRANS GROUP LPF E
NOISE GENERATOR
THROUGH DIR FILTER IN GROUP REC UNIT 40-56 KC (OR 60-76 KC)
TO GROUP REC UNIT
HGT
LGT
256 KC
236 KC
3700 CYCLES
HGR
LGR
TO GROUP REC UNIT
TO KEYING CIRCUITS

Figure 9 (left). Twin-channel unit. The slide visible at the top matches a slot in the mounting, Figure 7

Figure 10 (above). Diagrams of OB group transmitting and group oscillator units

minal and they are identical except for the crystal that determines the frequency of the oscillator and the carrier pick-off filter. The oscillator supplies the carrier for two channel modulators as well as the reinserted carrier that is transmitted over the line.

On the receiving side of the twin-channel unit, the 4-channel group comes from the group receiving unit and is amplified by the twin-channel regulator. The regulation is controlled by the particular carrier that passes through the selective carrier pickoff filter. The output of this regulator is applied to the channel band filters in two channel units. Each band filter selects the wanted side band before the signals are demodulated to voice. By maintaining a constant carrier power at the twin-channel unit output, each channel of a twin is regulated just as accurately as if each channel had an individual regulator. In this way the *O* system obtains performance equivalent to individual channel regulators with half the number of regulators.

A picture of the twin-channel unit is shown in Figure 9. The basic die-cast frame and general assembly are essentially the same for this unit, the group transmitting unit, and the group oscillator unit.

Figure 10 shows the connections and the functions located in the group transmitting unit and the group oscillator unit. The group transmitting unit contains a 2-stage feedback amplifier, with relatively flat transmission from 40 kc to 160 kc. It also contains a noise generator which amplifies noise from a gas tube to be inserted in the receiving group unit. There are two adjustments available in this unit, one to adjust the total power transmitted to the line, the other to adjust the level of the background noise introduced at this same terminal.

The adjustability of the transmitted level is highly desirable in the co-ordination of systems along an open-wire line. If a system is to be connected onto an open-wire line at a point which is not a repeater or terminal point for other systems on the same line, it is necessary to reduce the transmitted level so that it is about the same as the level of the other systems at this point.

There are three independent oscillators in the group oscillator unit. One oscillator supplies the 3,700-cycle signaling tone for the four channels. This oscillator has a Weinbridge circuit with a thermistor in the bridge to control the balance. An adjustment also is provided to set the magnitude of the signaling tone to the desired value. The other two oscillators are electron-coupled crystal oscillators to supply the group carriers. These carriers are used to group modulate the transmitted signal from the 180- to 196-kc band down to the desired line frequency band, and to group modulate the signals received from the line back up to the 180- to 196-kc band. There are eight different group carriers required for the four systems as shown in Figure 2.

Most of the differences between *OB*, *OC*, and *OD* systems are confined to the group receiving unit. A block diagram of

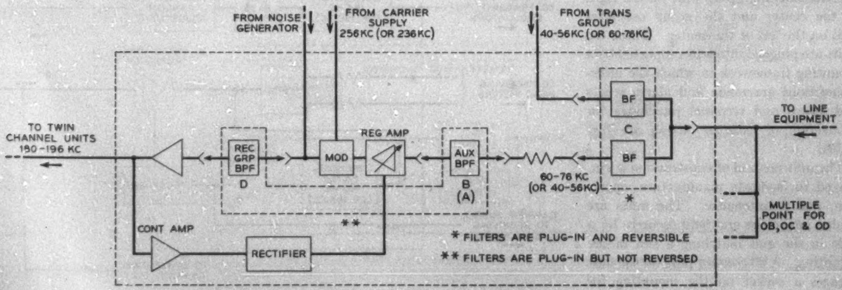
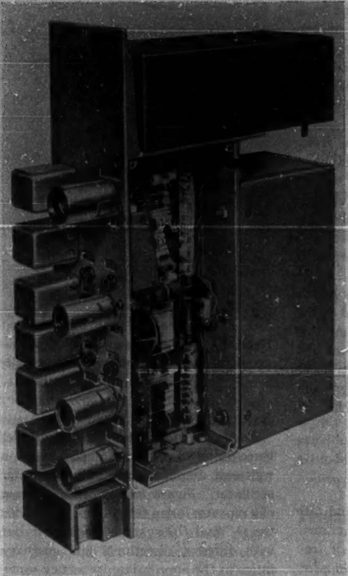

FROM NOISE GENERATOR
FROM CARRIER SUPPLY 256KC (OR 236KC)
FROM TRANS GROUP 40-56KC (OR 60-76KC)
TO TWIN CHANNEL UNITS 180-196 KC
REC GRP BPF
D
MOD
REG AMP
AUX BPF
B
(A)
60-76 KC (OR 40-56KC)
BF
C
BF
*
TO LINE EQUIPMENT
MULTIPLE POINT FOR OB,OC & OD
CONT AMP
RECTIFIER
**
* FILTERS ARE PLUG-IN AND REVERSIBLE
** FILTERS ARE PLUG-IN BUT NOT REVERSED

Figure 11. Diagram of OB group receiving unit



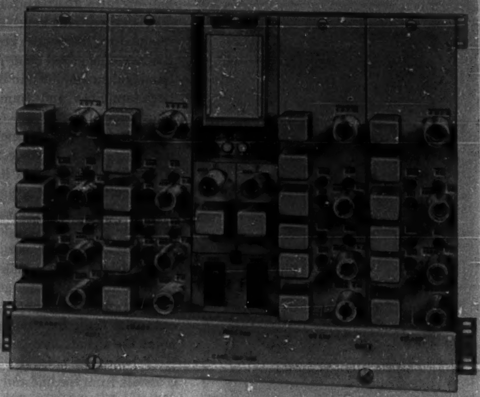


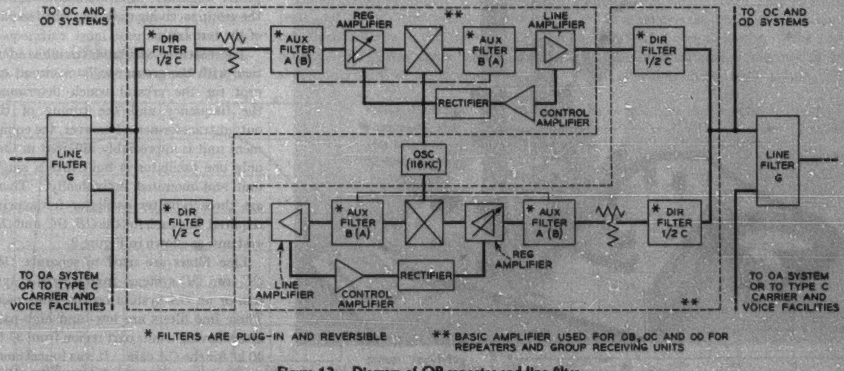
Figure 12 (left). Group receiving unit or repeater amplifier unit. Plug-in filters are at top and rear

Figure 14 (above). O carrier repeaters. Repeater 1 consists of the two left-hand units and the oscillator at the left center of the fuse panel

this unit is shown in Figure 11 and consists of a basic subassembly and two removable filters. One of these filters is the directional filter, the other filter is associated with the group modulator and contains both the receiving group band-pass filter and the auxiliary band-pass filter. The basic subassembly with its 2-stage amplifier, group modulator, and regulator circuit is built to give satisfactory operation for *OB*, *OC*, and *OD* systems. The particular system is determined by the filters plugged into the basic subassembly. This unit is shown in Figure 12.

By arranging the filters as plug-in components, the basic subassembly is used in 18 different situations. Six of these are at the terminal; low or high group receiving unit for the *OB*, *OC*, and *OD* systems. The other twelve situations are at a repeater, a low-high east-to-west repeater, or a high-low east-to-west repeater for the *OB*, *OC*, and *OD* systems and also for the corresponding west-to-east side of a repeater.

All the signals from the line filters are impressed on the input to the group receiving unit. The desired 16-kc band of frequencies is selected by the directional filter and passed through the auxiliary band-pass filter to the input of the 2-stage vacuum-tube regulator. The bias on the regulator tubes depends upon the


TO OC AND OD SYSTEMS
* DIR FILTER 1/2 C
* AUX FILTER A (B)
REG AMPLIFIER
**
LINE AMPLIFIER
* AUX FILTER B (A)
* DIR FILTER 1/2 C
TO OC AND OD SYSTEMS
RECTIFIER
CONTROL AMPLIFIER
LINE FILTER G
OSC (116KC)
LINE FILTER G
* DIR FILTER 1/2 C
* AUX FILTER B (A)
* AUX FILTER A (B)
* DIR FILTER 1/2 C
RECTIFIER
REG AMPLIFIER
LINE AMPLIFIER
CONTROL AMPLIFIER
TO OA SYSTEM OR TO TYPE C CARRIER AND VOICE FACILITIES
**
TO OA SYSTEM OR TO TYPE C CARRIER AND VOICE FACILITIES

* FILTERS ARE PLUG-IN AND REVERSIBLE

** BASIC AMPLIFIER USED FOR OB, OC AND OD FOR REPEATERS AND GROUP RECEIVING UNITS

Figure 13. Diagram of OB repeater and line filter

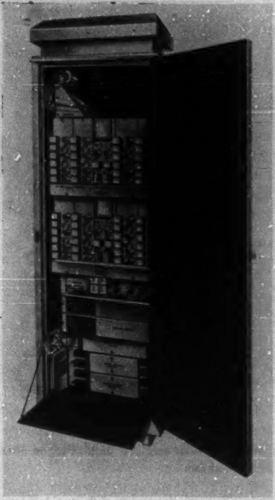


Figure 15. Pole-mounted cabinet. Containing four repeaters, line terminating panels, and rectifiers for operation from commercial a-c supply

magnitude of the output signal. This bias is obtained from the control amplifier and rectifier circuit. There is sufficient stiffness in this AGC circuit so that for a 40-decibel change in input signal the output will change only 1.5 decibel.

The modulator follows the regulator and operates at practically a fixed level. The desired side-band output of this modulator (180 to 196 kc) is selected by the receiving group band-pass filter. This band is then amplified in the 2-stage feedback amplifier and transmitted to the twin-channel units.

The only field adjustment in the group receiving unit is in the control amplifier. This adjustment makes it possible to set the magnitude of the group receiving unit output anywhere between 10 decibels below 1 milliwatt and 15 decibels above 1 milliwatt, where the normal adjustment is 10 decibels above 1 milliwatt.

Noise from the noise generator in the group transmitting unit is inserted at the output of the group modulator. At this point the signals have almost a constant level, hence the noise can be set to mask crosstalk from the line and the magnitude of the noise is practically independent of variations in line loss.

Repeaters

The repeater of the *OB*, *OC*, and *OD* systems performs four basic functions. It separates out the two groups of frequencies used for the two directions of transmission in its particular system, translates and inverts each incoming group by modulation to the opposite group, amplifies the signals, and automatically regulates the repeater gain to compensate for changes in line loss. The *OA* repeater performs three of these functions, omitting the modulation function since frogging is not employed in the *OA* system.

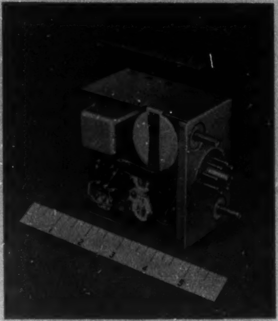


Figure 17. Channel band filter. Cover removed to show crystal and ferrite core inductor

Each *OB*, *OC*, or *OD* repeater includes three units, an east-west amplifier unit, a west-east amplifier unit, and a repeater oscillator. Figure 13 is a block diagram of a repeater for an *OB* system. Those for the *OC* and *OD* systems are similar but with different directional and auxiliary filters in the appropriate frequency range and with different oscillator frequencies.

In these systems employing frogging, alternate repeaters along the line are of two types, high-low and low-high, named for their input and output groups. The repeater amplifier units in these two repeater types are identical. To change a repeater from a high-low to a low-high, the directional filter and the auxiliary filter are rotated 180 degrees and placed back in their mountings. The *OA* repeater is like these but with modulation and the oscillator omitted and amplifiers designed to operate at lower frequencies. The repeater amplifier units are identical with the group receiving units described as part of the terminals.

The repeater oscillator circuit is identical with the group oscillator circuit except for the crystal which determines the frequency and the tuning of the output transformer. However, the equipment unit is appreciably different in that only one oscillator is built into a single unit and mounted individually. There are three repeater oscillator frequencies required, one each for the *OB*, *OC*, and *OD* systems, as shown in Figure 2.

Line filters are used to separate *OB*, *OC*, and *OD* systems from a type-*C* system or an *OA* system on the same pair. These line filters are low- and high-pass filters having a cutapart region from 36 to 40 kc for the *OA* case. It was found more economical to separate *OA* from the other

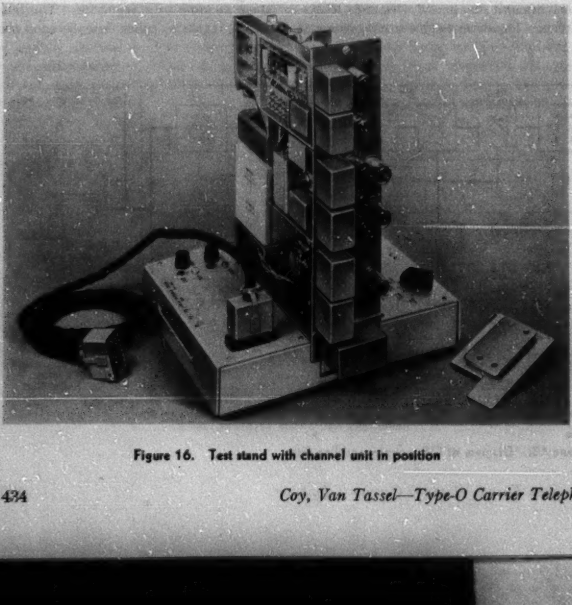


Figure 16. Test stand with channel unit in position

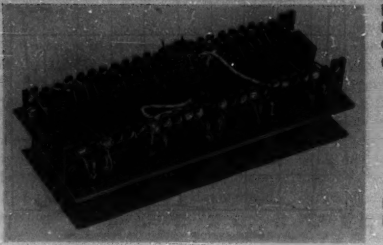


Figure 18 (left). Directional filter. Cover removed to show ferrite core inductors

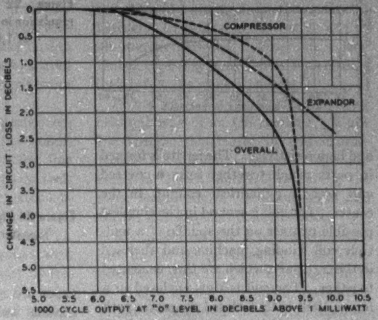


Figure 20 (right). Limiting load capacity

3 systems by means of high- and low-pass filters than to continue the band-pass filter method used to separate *OB*, *OC*, and *OD*.

The repeater assembly, shown in Figure 14, is similar to that of the terminal, but five units mounted on one shelf constitute two repeaters. In the illustration the two left-hand units are the west-east and east-west repeater amplifier units of repeater 1, the two right-hand units the corresponding parts of repeater 2, and the center unit a mounting for the two repeater oscillator units, the alarm relays and lamps and the fuses for the two repeaters. The center unit as well as the repeater amplifier units are connected to the mounting shelf by plugs and jacks so that all are removable for maintenance and convenient for bench assembly in manufacture The complete repeater mounting, like the terminal mounting, is shop wired thereby simplifying the installation to that of mounting it on the rack and connecting the external wires to the terminal strip.

At remote locations on light routes, repeaters are mounted in cabinets placed on a line pole or on a crossarm between a line pole and a stub pole. One cabinet, Figure 15, contains four repeaters with rectifiers and power control equipment for operation from commercial a-c supply. A similar cabinet when desired contains batteries, dynamotors, and charging equipment for reserve in the event of power failure. A small 24-volt battery with a dynamotor furnishes the 130-volt supply for two repeaters. An alarm is provided over one of the line wires to an attended point to indicate power or fuse failure.

Testing

As an integral part of any system provision must be made for testing its performance and for trouble location. Pin jacks are mounted in the face of the plug-in units and connected to the circuit at suitable points, for testing. These tests are made with a vacuum-tube voltmeter and are designed to show which unit is in trouble. There is no attempt here to determine what is wrong with the plug-in unit. When the bad unit has been located a good spare unit is substituted and the unit in trouble is removed to a convenient repair center where more detailed and complete maintenance can be performed on it.

Under certain conditions more complete tests and adjustments which require access to the interior of the units may be required. For these cases a test, stand shown in Figure 16, is provided. A unit may be withdrawn from the mounting on the bay, plugged into the test stand, and connected by a cord to the jack from which it was removed. With the unit effectively connected back into the system, either local or over-all system tests can be made on the unit.

Filters

The close frequency grouping of *O* carrier channels is made possible and economical by the combination of:

1. Twin channel operation with a new dissymmetrical band filter in which a crystal and a ferrite core inductor are combined to give a flat passband and a discrimination characteristic steep on one side.
2. Group filters which pass a band 16 kc wide with adequate suppression against an adjacent band only 4 kc away with opposite directions of transmission, made practicable by means of ferrite core inductors.

The channel band filter, Figure 17, is a very compact unit employing a crystal and a ferrite core inductor. The directional and auxiliary filters, such as Figure 18, have no crystals but contain as many as ten ferrite core inductors. The cores of these inductors are a mixture of ferrites providing both high permeability

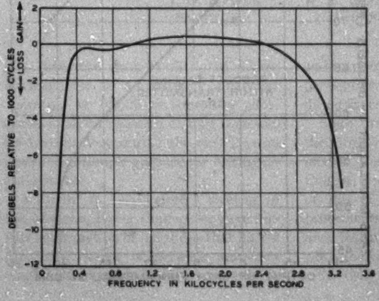


Figure 19 (left). Frequency characteristic of a typical message channel

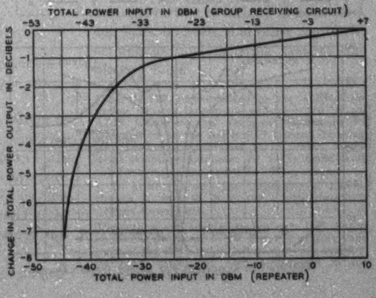


Figure 21 (right). Repeater and group receiving circuit regulation

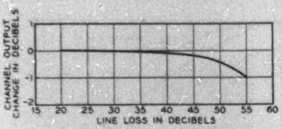


Figure 22 (left). Over-all regulation in a message channel at 1,000 cycles

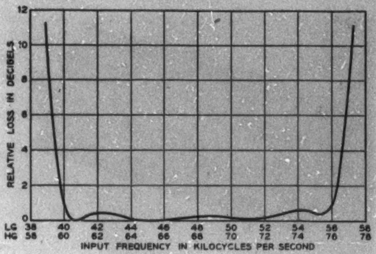


Figure 24 (right). Typical OB repeater characteristic

and high resistivity. The ferrite is formed in parts which together form a covered cup with a cylindrical column in the center. The coil is wound in the simplest possible manner on the spindle of a multiple coil winding machine and enclosed in the core parts which are then cemented together. The coil is thus enclosed in a ferrite shell with a closed magnetic path so that crosstalk to associated circuits is at a minimum. An inductance variation sufficient for manufacturing adjustment of the filter is obtained by mechanically changing the air gap in the core. A coil occupying less than a $1^1/_2$-inch cube can be made for use at carrier frequencies having a "quality factor" Q well over 500. The winding, being formed before assembly, is economical to make and assemble. Ferrite is a very permanent material so that it is not affected by humidity or subject to aging.

Universality of Parts

One objective in designing the type-*O* system has been to build up the terminals and repeaters for the four systems by flexible combinations of the fewest practicable units and subassemblies. This is of obvious advantage both in manufacture and in maintenance.

By generating all eight of the 4-channel groups indicated in Figure 1 in the common range of 180 to 196 kc, only four channel-unit types and two twin-channel types are needed for all systems. Each channel unit is composed of three subassemblies. Of these the compressor subassembly and the expandor and signaling subassembly are identical not only in all *O*-systems channels but also in the 12 channels of the type-*N* system. The carrier frequency subassembly is the same for all *O* channels except for a plug-in band filter. This filter is made in two types each serving two of the four channel units by reversal in its socket. There are two types of twin-channel units which differ only in the crystal that determines the oscillator frequency and the pickoff filter that selects the incoming carrier. The twin-channel units are interchanged in position in the terminal mounting framework to change a high group transmitting terminal to a low group transmitting terminal.

The group transmitting unit is universal for all *OB*, *OC*, and *OD* systems. The group receiving unit is adapted to *OB*, *OC*, and *OD* high group transmitting or low group transmitting terminals by plugging in filters of appropriate design. It is adaptable in the same way to use as the east-to-west or west-to-east unit of a repeater. Units for the *OA* system are somewhat different due to its low frequency range. The same directional filter is used at terminals and repeaters and is adapted to high group input or low group input by rotating it 180 degrees in its socket. The group oscillators and repeater oscillators for the four systems differ only in the crystals.

Identical die castings are used for the terminal and repeater mounting shelves. The castings for the twin-channel units, group transmitting unit, and group oscillator are the same except for the faces. They can all be cast in the same die by changing the slide which forms the face, a relatively inexpensive arrangement. In this way the total demand for castings is satisfied by relatively few designs, each made in larger number. The cost of making dies is reduced severalfold.

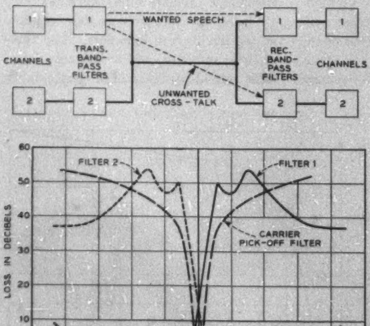


Figure 23 (left). Suppression of interchannel crosstalk

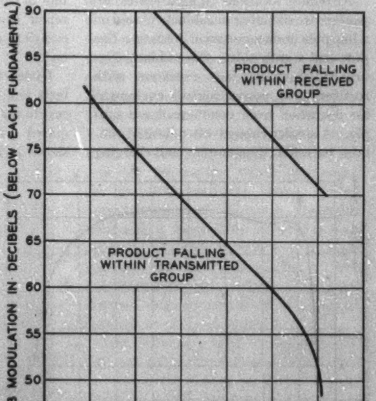


Figure 25 (right). Third order modulation in an OB repeater

System Performance

A measure of the over-all performance of the type-*O* system is shown by the typical characteristics. Figure 19 is a typical over-all frequency characteristic of the message channel. In the normal use in the telephone plant more than one type-*O* system may be used in a particular built-up circuit. A built-up connection that contains four type-*O* systems would attenuate 250 cycles about 10 decibels, relative to the loss at 1,000 cycles and on the high frequency end 3,000 cycles will be attenuated about 10 decibels.

All carrier systems have a finite limit in the peak power that can be transmitted. The extent to which the power peaks are limited by the *O* system is shown in Figure 20. Normally a *O VU* talker will have a peak 10 decibels above the rms value about 1/10 of 1 per cent of the time. This is about the limiting peak for this system.

During normal operation with dry weather conditions along the open-wire line and no "static" fields due to storm conditions, the line noise and repeater noise is negligible. The background noise, necessary to mask crosstalk from other paralleling systems, is introduced at the receiving terminal and is about 57 decibels below a loud or *O VU* talker. However, during periods with thunderstorms along the open-wire line the noise on these systems will be primarily "static" as picked up on the open-wire pairs. The increase in noise may be 5 to 15 decibels and on rare occasions even more, dependent upon the severity and location of the storm.

The loss in an individual repeater section depends upon the weather conditions, and may change from a dry line loss of 13 decibels to a loss of 50 decibels during sleet conditions. The extent to which a repeater compensates for the change in flat loss is shown in Figure 21. That part of the loss that varies with frequency is compensated in part by frequency frogging and in part by the twin-channel regulator. The over-all change in the 1,000-cycle net loss for a 35-decibel change in line loss is shown in Figure 22.

In most carrier systems the extent to which adjacent channels are separated is measured in terms of crosstalk from one channel into the other. One path in the type-*O* system by which energy from channel 1 of Figure 23 is able to be heard in channel 2 is through the channel band filters. At the transmitting end the unwanted side band of channel 1 is attenuated by filter 1 to give a clear space for the wanted side band of channel 2. Then at the receiving end the desired speech energy from channel 1 is applied to both channel 1 and channel 2 receiving band filters. In this case, filter 2 must suppress the channel 1 energy from the channel 2 listener. In general the crosstalk due to all sources between channels in a particular type-*O* system is more than 60 decibels below the wanted speech.

Figure 24 is a typical gain frequency characteristic of an *OB* repeater. While the specific departures in this characteristic from a flat value are not duplicated in the *OC* and *OD* repeaters, the same magnitude of variation exists and the typical shape is about the same.

A measure of the interchannel crosstalk due to the nonlinearity of the components is shown in Figure 25, in the form of a plot of the 3rd order modulation products produced in a line repeater. In this instance the line amplifier is not the chief source of this modulation. The ferrite inductors in the group filters, the group modulator, and the regulating amplifier introduce more distortion than the line amplifier. Taking into account the 28-decibels compandor advantage, the over-all modulation crosstalk is more than 60 decibels below the normal talkers.

From these characteristics it is apparent that good service standards have been maintained and not sacrificed for low cost.

Reference

1. A NEW TELEPHONE CARRIER SYSTEM FOR MEDIUM-HAUL CIRCUITS, R. S. Caruthers, H. R. Huntley, W. E. Kahl, Ludwig Pedersen. *Electrical Engineering*, volume 70, August 1951, pages 692–97.

No Discussion

Design Improvements and Operating Experience with 10-Kc Network Analyzers

J. D. RYDER
FELLOW AIEE

W. B. BOAST
MEMBER AIEE

IN 1946, the authors disclosed a new design for an a-c network analyzer[1] operating at 10,000 cycles. The pilot-model analyzer reported at that time was used extensively during the next 2 years upon small sample systems for testing its performance.

A report[2] in 1949 before a meeting of the Midwest Power Conference gave further confirmation of the thesis that higher frequency design of network analyzers employing complete electronic generation and metering was feasible and practical—particularly from the advantage of reduced investment costs.

Since the first report on the design of the 10,000-cycle a-c network analyzer in 1946, work has continued at Iowa State College, and since 1949 at the University of Illinois, looking toward further improvements in both economic and technical features of such high-frequency network analyzers.

In 1949 the construction of a 16-generator analyzer was begun at Iowa State College with the co-operation of various utility and consulting firms in the Iowa area. This analyzer was placed in service in September 1950, with 100 lines and 50 loads, and in addition to its use for research and instructional purposes by Iowa State College, was operated commercially for 28 weeks during the past year. This analyzer, shown in Figure 1, is now being enlarged further to include 16 generators, 140 lines, 66 loads, and other accessories as well as dual metering systems and plotting tables.

Paper 52-310, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted December 26, 1951; made available for printing September 3, 1952.

J. D. RYDER is with the University of Illinois, Urbana, Ill., and W. B. BOAST is with Iowa State College, Ames, Iowa.

The authors are indebted to Professor J. E. Lagerstrom and Professor D. D. Robb, operating engineers of the Iowa State College network analyzer, for their suggestions and co-operation in the preparation of this paper.

Figure 1. The Iowa State College 10-kc network analyzer

Since 1949, an analyzer has been built at the University of Illinois, shown in Figure 2 having 20 generators, 152 lines, 56 loads, and accessories. This latter unit also includes two instrument desks and a plotting table.

Construction of these two large-scale 10,000-cycle analyzers has borne out the predictions made in 1946 concerning lowered costs inherent in the choice of the higher operating frequency, and further technical improvements have been achieved because of the original choices of frequency and the per-unit values of 10 volts, 0.1 ampere, 1 watt, and 100-ohms base impedance. These values have permitted taking advantage of modern electronic techniques and equipment and circuit advances.

Description of Units

Line and load units have been redesigned, resulting in an appearance as in Figure 3 for the line units and as in Figure 4 for the load units. These incorporate inductors having molybdenum-permalloy dust cores of constant permeability over the current ranges to be expected in normal operation. A third and lower decade has been added to each of the series resistance and series inductive reactance of the line pi sections. The lower inductance decade, having a total reactance of 1 ohm, utilizes a design of a tapped molybdenum-permalloy-core reactor having half turn steps. These were achieved by winding on two cores, the turns sometimes linking both, sometimes linking only one of the cores, thus achieving satisfactory accuracy of tap inductance. The line units were reduced in size in the Illinois design, thus saving a number of square feet of panel space, and the analyzer at the University of Illinois has been built in the same number of lineal feet of panel as the slightly smaller unit at Iowa State College. Load voltmeters have been added with normal indexes to indicate voltage present on any load. In addition, switches have been added at both ends of the line unit and at the input of the load in the Illinois design to simulate circuit breakers, permitting ready opening of the network at any of these points. This eliminates the necessity of pulling plugs at the patch panels to drop loads or disconnect portions of a system

Through the use of the molybdenum-permalloy coils, the Q, or reactance to resistance ratio of these units has been raised above 100 in all cases. Thus the resistance of the inductors in the circuit can be neglected at all times, since the error due to the resistance of the inductive elements always will be less than 1 per cent of the reactance value.

Isolation of capacitive susceptance of the pi-section line units from chassis ground was not possible on the early pilot model. A dual jack system has been incorporated into the ground connection to permit establishment of an isolated zero-potential bus. This feature has been found necessary where capacitive susceptance is set in sequence networks isolated from the positive-sequence zero-potential bus, which is generally established as chassis ground on the network analyzer.

The generators are essentially electronic feedback amplifiers, operating from a single source at 10,000 cycles. Output power has been increased by improved design and use of greater amounts of feedback, and currents of 500 per cent of base (5.0 per unit) can be obtained. The generator output voltage is variable over the range of 0.3 to 1.5 times base value and is regulated and constant at the selected value up to load current magnitudes exceeding 500 per cent or 5.0 per unit, for all types of load circuits, leading, lagging, or resistive. Thus the generator impedance can be considered zero for all loads, and network changes made without need for readjustment of generator voltage.

An electronic phase-shifting circuit has been developed which permits 360 degrees of phase rotation of the output voltage of a generator on one dial. Gen-

Figure 2. A portion of the new 10-kc network analyzer at the University of Illinois

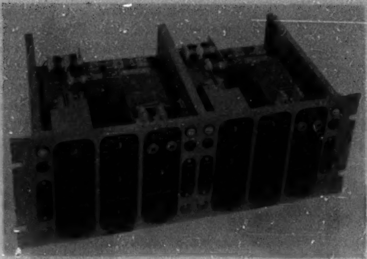


Figure 3 (left). Assembly and panel for two line units, showing circuit opening switches and auxiliary bus jacks

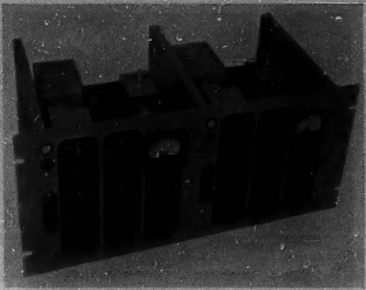


Figure 4 (right). Assembly and panel for two load units

erator panel construction for the Illinois development is indicated in Figure 5, which shows at the top four sets of generator instruments, indicating generator volts, watts, and vars. The white dial areas will be directly calibrated for voltage and phase angle, each white area carrying the two controls for one generator. Vernier adjustments are provided on both voltage and phase output of the generator. The tap switches for inserting generator reactances are located on the panels below the generator dials.

The generator voltmeters are rectifier operated through isolating electronic circuits, while the watt- and var-indicating instruments employ electronic circuits using thermal converter principles.

Instrumentation

The metering system is the heart of a network analyzer. The original plan of direct-reading instruments for magnitudes and phase of both voltage and current as well as instruments for direct indication of watts and vars in the original pilot model has been followed in the full-scale analyzers previously described.

Separate voltage and current preamplifiers, each controlled closely for strict magnitude fidelity and no phase distortion provide signals proportional to voltage and current. These signals in turn are applied to meter amplifiers for the previously listed indicating instruments.

The full-scale ranges of voltage metering employed on the pilot model were 0.3 and 1.5 per unit. Several changes have been made upon these ranges, and the presently established ranges are 0.3, 1.5, and 3.0 per unit.

The full-scale ranges of current metering employed on the pilot model were 0.6 and 2.0 per unit. These were found to cover an inadequate range of system possibilities, and several changes in range have been made since that time. The present current metering ranges are 0.2, 1.0, and 5.0 per unit.

Further effective change in the direction of expansion of the current range is being planned, but not as a direct modification of full-scale currents. Methods are being studied for simultaneously modifying all generator voltages by a constant factor. This will effectively increase the maximum current metering possibilities for short-circuit studies when the fault point is close to large generating capacities.

A view of the metering desk and control console is shown in Figure 1 for the Iowa State College design and for the Illinois design in Figure 2. The desk with its attached table becomes a complete data recording center. The instruments comprise voltmeter and voltage phase meters, current and current phase meters, and watt and var meters. Metering of voltage and current on a rectangular component basis also is available as desired in both the Illinois and Iowa State design. Instruments may be inserted at will in the network by pressing three numbered buttons corresponding to the assigned number of the unit to be metered, circuit selection being by relays.

In addition, the installations being described are provided with a duplicate metering desk and a plotting table. Views of the plotting table at the University of Illinois are shown in Figures 6 and 7, the latter being a rear view, showing the small banana jacks by which lights under the plotting surface can be connected by cords and plugs to the desired units in the analyzer. The working surface of the plotting table, Figure 6, is a translucent sheet behind which are positioned 1,452 telephone switchboard lamps at 1-inch intervals horizontally and 1/2-inch intervals vertically. This plotting table will become an integral part of the second instrument desk, permitting data to be recorded directly on system maps.

It should be pointed out that these desks are capable of reading electrical quantities at any point in the whole analyzer, and thus work may go on with one large problem being metered simultaneously at two desks, or the network analyzer may be broken into two parts, and operation performed on two systems simultaneously. No error will be introduced should both desks be connected to read the same electrical quantity simultaneously.

Effect of Leads and Shunt Capacity

The original design of these analyzers was based on the requirement that the series impedances of leads from the respective units to the patch panels should not in any case introduce an error exceeding 0.005 per unit or 1/2 per cent of the base values. Average impedance values measured for these leads on the analyzer at Iowa State College are given as follows.

It should be noted that the shunt values are taken with the tuning apparatus in place.

Average leads—units to patch panel
Series: $R=0.005$ per unit or 0.5 per cent
$X=0.004$ per unit or 0.4 per cent
Shunt: $g=0.0009$ per unit or 0.09 per cent
$b=$ essentially zero

These values, as well as complete shielding, are achieved through the use of large-size twin-conductor coaxial or "Twinax" cable for these leads. While all units are completely shielded, it has been found that the use of the low 100-ohm base makes shielding relatively unimportant, either against magnetic or electric pick-up.

Allowance was made in the analyzer design at 10,000 cycles for the effect of shunt capacitance from these leads to ground. This capacitance plus an additional adjustable capacitance is made parallel resonant and resistive by shunting with a high-Q inductance coil. The effect of

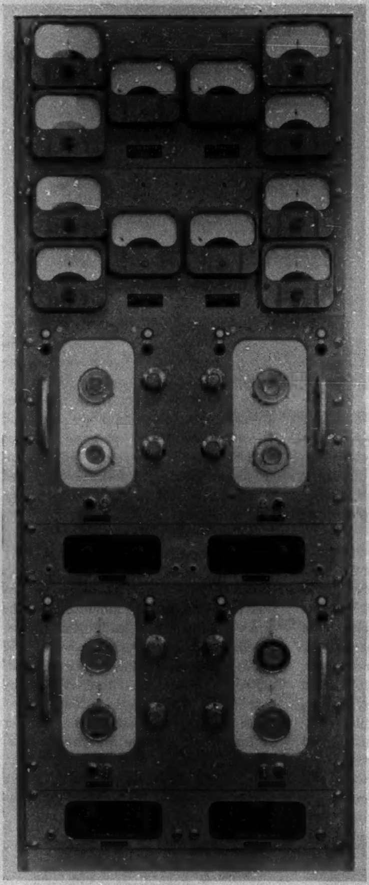


Figure 5 (left). View of generator panel, showing instruments, generator controls, and generator reactance switches for four generators of the University of Illinois design

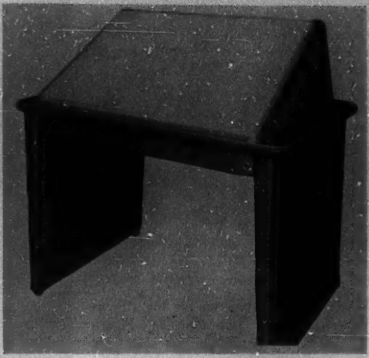


Figure 6 (right). Front view of the plotting table at the University of Illinois

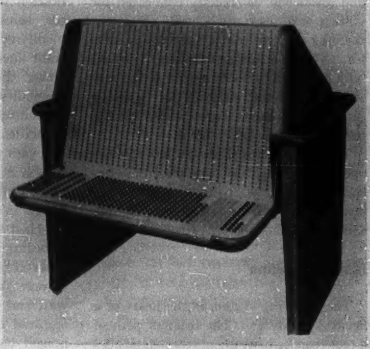


Figure 7 (right). Rear view of the plotting table of Figure 6, showing arrangements for plug connections

parallel lead capacity and unit capacity to ground then appears as part of a parallel resonant circuit with a resistive shunt impedance exceeding 110,000 ohms in all cases, and its effect is thus completely negligible when in parallel with a few hundred or a few thousand ohms, which are usual load values.

The metering bus, which is inserted into the network by relay operation to convey the various electrical quantities to the master shunt and meters, can be sectionalized by relays and each section individually shunt-tuned to eliminate the effect of parallel capacity. This bus is made of close-spaced 1/4-inch copper tubing and residual impedances are given here.

Series: (Average length)	$R=0.0005$ per unit or 0.05 per cent $X=0.002$ per unit or 0.2 per cent
Shunt: (total bus)	$g=0.0002$ per unit or 0.02 per cent $b=0.0000$ per unit or 0.00 per cent

The series impedance of the bus together with the 0.1-per-cent impedance shunt of the current metering system presents an insertion impedance of 0.25 per cent. This impedance can be compensated in each individual unit if desired, so that the meter-bus insertion error is made zero for any impedance circuit larger than the sum of the residual impedances of the leads and of the meter bus.

Also associated with the metering system has been the problem of capacitive coupling of various parts of a network through the capacitance of the metering relays connected between all units and the meter-bus system. This problem was studied carefully before the pilot model was established and was one factor entering into the choice of the relatively low impedance base of 100 ohms. Experience with the Iowa State College 100-line 50-load unit analyzer indicates that a predicted shunt reactance of the order of 10,000 ohms does occur for full network connections. Under the worst combination of network arrangements, this reactance appears across a series network impedance of not more than 100 ohms and usually much less. The error, therefore, may be of the order of 1 per cent. The error is detectable by checking voltages over the system as the metering system is operated. A system of meter-bus sectionizing relays would raise the undesired shunt reactance to the order of 100,000 to 300,000 ohms depending upon the location in the network that the meter bus is inserted. This is a refinement that was contemplated before the Iowa State College analyzer was constructed and control contacts on the relays in the units were provided for such use.

Transformers

It has been found possible to wind transformers on Hipersil cores or Conpernik rings and to achieve series and shunt impedance values which in most cases permit neglecting the presence of the transformer in the circuit. As an example, tapped autotransformers, for insertion where needed in the network, have shunt impedances exceeding 55,000 ohms (550 per unit or 55,000 per cent). Series resistance and reactance approximate 0.2 and 0.1 per cent of base impedance respectively at the normally used taps.

The isolating unity-ratio transformers of the pilot model analyzer were built in a single design favorable to low-voltage high-current applications. Experience showed that the shunt input impedance of the unit was relatively low for normal voltage isolation applications. The design of this transformer was revised so that it now can be used for either low-voltage high-current use or for high-voltage low-current applications. The transformer is a 4-winding design with switching arranged so that the windings can be connected by pairs either in parallel or in series. The resulting transformers have coefficients of coupling exceeding 0.9999. Thus, transformers can be inserted in circuits at will and considered as pure voltage or current-ratio devices without regard to changes in circuit impedance introduced by the transformer losses or magnetizing currents.

Additional Features

In addition to the somewhat standardized units already in service, work is either completed or in progress on a number of additional features, which in most cases are made possible by reason of the choice of base impedance and power and the higher operating frequency. These features include an electronic phase-shifting transformer for insertion where desired in the network. This device, developed as a graduate research project at Iowa State College, will introduce any desired phase shift up to 15 degrees, between primary and secondary quantities, at all load values without introduction of internal losses and with no laborious trial and error adjustments as by the previous "load-generator" method.

A constant kilovolt-ampere load also has been designed at the University of Illinois. Its impedance varies in magnitude as the square of the applied voltage.

Servo control of the generators from net system accelerating power also will permit automatic determination of system swing curves.

Equipment for synthesizing symmetrical component values from a 2-component analyzer study has been developed. As soon as the second metering system is completed, this development will be added to the Iowa State College analyzer. Its use will remove the need for laborious manual synthesis of line current values from the symmetrical component values as obtained from analyzer studies.

Accuracy Checks

Two general types of checks can be made on the various types of system studies. An absolute check is possible only on relatively simple systems that can be calculated for complete performance. Several such checks were run on the pilot model.[2]

For complex systems relative checks are the only type that can be made with certainty. A partial, absolute check can be made against actual system-load data, but variance of readings with these data is not always the fault of the network analyzer.

One rather complete relative check on a system-load study was made on the Iowa State College analyzer during the Spring of 1950. The system involved 5 generator busses feeding into a network consisting of 56 lines, 38 loads, 5 tap ratio transformers and 2 bus capacitors. Line unit losses in watts and net vars were computed from average line voltage and current readings and the line parameters. The percentage component watt and var errors based upon total volt-amperes of the system were found to be 1.36 per cent and 2.98 per cent respectively. The higher error on vars is probably attributed to the shunting capacitive reactance of the relays of the meter bus since the Iowa State analyzer has been operating without a sectionalized meter bus system. A reduction in such errors is anticipated when the sectionalization feature is completed.

Checks at busses constitute another form of relative check. There is a slight tendency for the busses to apparently generate a small amount of watts and vars. One reason for the apparent generation of vars is the manner in which the inherent shunt capacitance of the line units is tuned for the elimination of undesired capacitive effects. The total compensation for the inherent capacitance is placed in the units and inside the metering relays. Thus, the total capacitance is balanced exactly. On the bus side of the metering points, however, untuned capacitance results; and inside the metering points, the circuit is overtuned inductively. An exact tuning by three parts would be possible one on each lead circuit outside the metering points and the third part inside the metering points. The complexity of this manner of tuning is not believed justifiable for the slightly better balance achieved.

It should be stressed that the total error caused by undesirable shunt capacitance effects is not reactive in nature, but is resistive. The use of relatively high-Q compensating inductors at the operating frequency leaves residual very high resistance apparent loads within each unit.

The system of inserting the meter bus in succeeding units tends to reduce the true current in each circuit as it is read. The error is quite negligible when metering circuits possessing any appreciable series impedance. The tendency in networks, however, is that circuits supplying watts and vars to a bus are often somewhat lower in impedance values than those receiving watts and vars from the same bus. The insertion error then is in such a direction as to cause the sum of metered values leaving a bus to be greater than the total being received. This tendency is observed to predominate in most studies. The effect is particularly noticeable, of course, when parallel, very low impedance circuits are metered. Effective compensation of this insertion effect is planned as an integral part of the new units of the expansion program at Iowa State College with addition of this feature to other units already constructed. This compensation will take the form of a matched impedance in each unit that will reside there permanently except when the meter bus is inserted.

As has been indicated, the Iowa State College analyzer has been operated during its first years with a single desk and metering system. Outages caused by failure of equipment were reasonably frequent during the first thousand hours on the "running-time" meter, but the continuity of operation since that initial period has been very good. Difficulties caused by cold-solder connections, plugging-cord short circuits, switch-contact resistance, tube failures, weak filter capacitors, and relay contacts accounted for most of the initial outages.

Plugging-cord short circuits have been eliminated by redesign of the original cords. Poor contacts have been isolated and replaced or repaired, and tube failures have been reduced by a revised schedule of their operation. Plate supply voltages are removed by time-clock operation at night and filaments are operated at reduced voltage. Filament voltage is raised to normal level at 7 a.m. and plate-supply voltage is applied about 30 seconds later. Spare amplifiers, frequency standards, and power supplies are an effective means of reducing time of outage when any of these parts fail. Monitor meters have been placed on the input and outputs of all preamplifiers as well as on supply-signal circuits and power supplies as a means of quickly isolating difficulties. Substitution of parts may then be accomplished quickly with maintenance work done carefully rather than hurriedly. Approximately 2,200 hours were recorded on the "running-time" meter for commercial operation on the Iowa State College analyzer during the first year. A total of only 8 hours of outages occurred during the last 1,200 hours of recorded operating time. We believe this can be reduced to essentially zero with spare parts as suggested previously.

Conclusion

Experience in the design and construction of three analyzers of the high-frequency electronic type has shown a definitely decreased cost as a major advantage for these devices. The design also makes possible additional features which should further improve the operating ease and reduce the time required for obtaining certain types of information by network analyzer. As a result of 5 years of operating experience with the high-frequency analyzers at Iowa State College, and the recent beginning of operation with the larger unit at the University of Illinois, it appears that the high-frequency analyzer has been definitely proved.

Because of the considerably lowered cost of the high-frequency analyzer, there is no economic pressure to show continuous 52 weeks' operation of an expensive facility in order to obtain a return on the investment. Thus, ample time is and should be available in these installations for research in connection with application of the network analyzer in other fields, and also for research work on improvements to the network analyzer itself. Both of these areas appear exceedingly fruitful.

References

1. A New Design for the A-C Network Analyzer, J. D. Ryder, W. B. Boast. *AIEE Transactions (Electrical Engineering)*, volume 65, October 1946, pages 674–80.

2. A Network Analyzer Operating at 10,000 Cycles, W. B. Boast, J. D. Ryder. *Proceedings*, Midwest Power Conference (Chicago, Ill.), volume 11, 1949, pages 72–78.

Discussion

R. B. Squires (Westinghouse Electric Corporation, East Pittsburgh, Pa.): The improvements described by the authors have eliminated most of the objections to the design given in their previous paper. No doubt further operating experience will prove the desirability of other changes such as a wider range of current scales on the master instruments.

Because these installations are the only full-scale electronic network calculators, it would interest many members of the Institute to know how many electron tubes are replaced in a year of commercial operation. Have any records been kept of this replacement?

Table I. Iowa State College Tube Replacement Record

Classification	Total Tubes in Service	Monthly Percentage Replacement
Power and signal supplies and regulators	64	2.8
Metering amplifiers	49	3.5
Generators	160	1.2

J. D. Ryder and W. B. Boast: The tube replacement record on the Iowa State College Analyzer is shown in Table I of the discussion. This represents 26 month's operating experience of tube replacements based upon periodic checks. In general, tubes are replaced when the tested mutual conductance falls below two-thirds of rate values.

The Single-Core Magnetic Amplifier as a Computer Element

R. A. RAMEY
MEMBER AIEE

THE magnetic amplifier and magnetic devices in general have garnered an enviable reputation for reliability and maintenance-free operation[1] in their not-so-short history.[2] For these reasons these devices have been a subject of considerable interest[3,4] to the computer field wherein electron tube failure is a major headache. There have been many magnetic devices developed for use in this field, but for the most part they either have been slow (requiring several cycles of excitation frequency for build-up of output) or have not exhibited any considerable power gain. Speed of response is quite important for this determines computing speed. Power gain is quite important, for in a given computer these devices must be cascaded so that one device operates others as well as associated registering or indicating devices.

The advent of a magnetic amplifier circuit with discrete and fast response time[5] logically leads one to the thought that such amplifiers may have application in the digital computer field. This paper is directed toward presenting a few of these possible applications with experimental illustration of a flip-flop, a coincidence counter, and a ring counter. Application of basic devices of these types in computer circuits is well known and varies according to the problem involved. This paper, however, will try to cover only these basic magnetic elements and leave specific application to the imagination of the computer engineer.

Single-Core Amplifier

Two versions of the single-core amplifier are shown schematically in Figure 1. Each consists of a magnetic core coupling two circuits via one or more

---

Paper 52-293, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, New Orleans, La., October 13–17, 1952. Manuscript submitted June 16, 1952; made available for printing September 4, 1952.

R. A. Ramey is with the Naval Research Laboratory, Washington, D. C.

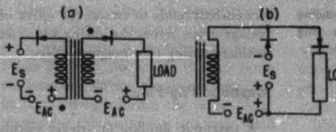


Figure 1 (left). Single-core magnetic amplifier

(A) Double winding
(B) Single winding

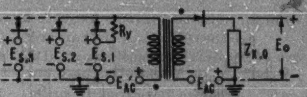


Figure 2 (right). Magnetic coincidence counter

windings. The load circuit consists of a source of alternating voltage E_{ac}, a rectifier, and a load in series with one winding, and the control circuit consists of a source of alternating voltage E_{ac}', a rectifier, and a control voltage in series with the second (or same) winding of the core. E_{ac} and E_{ac}' have the same frequency and phase. They have magnitudes proportional to the turns ratio of the saturable transformer and have the instantaneous polarities noted. The dots (Figure 2) refer to winding polarity. Since the 2-winding version offers a greater measure of versatility, it is considered henceforth in this article. In less demanding designs, however, the single-winding amplifier offers less expensive construction.

The operation of this amplifier can be considered in two separate periods:[4] reset and gating. During the half-cycle with polarity, as indicated in Figure 1, current can flow only in the control or reset circuit because of the direction of the rectifiers. This causes the magnetic core to proceed from saturation (reset) an amount dependent upon E_s and E_{ac}'. During the next half-cycle current can flow only in the output or gating circuit causing the magnetic core to proceed to saturation (gate), thereupon allowing the alternating voltage E_{ac} to be impressed across the load for a period determined by the degree of reset of the previous half-cycle. In usual amplifier practice, the output current or voltage integral is a relatively linear function of the control voltage; however, in the circuits to be considered herein, the control voltage E_s will be either zero or else large enough to prevent reset. The results in the output circuit will be either zero or full output.

This amplifier is characterized by fast response (output in any half-cycle proportional to the control voltage of previous half-cycle) and high gain. There is no exponential-like rise of output, but rather full output is attained without an intermediate state, after a discrete delay time of one-half cycle of excitation frequency.

Sudden changes of state are quite easily attained electronically with practically no delay time, but ordinarily magnetic devices exhibiting gain also exhibit exponential-like response characteristics. Under optimum conditions a power gain of 100 may be attained in a single stage in conventional circuitry with response time (63 per cent of final output) of 1 or 2 cycles of power frequency. In computing devices, wherein a 100 per cent change in state is desired, one would have to wait approximately 3 to 6 cycles for full output to be approached. The ready solution to this difficulty is an increase in the excitation frequency to some 10 to 20 times the computer repetition rate. Although magnetic devices can be made to operate at high frequency, the problem, as with most other devices, is seriously complicated as the frequency is increased by even one order of magnitude.

By making use of the rapid response of the single-core amplifier and meeting the condition imposed by the half-cycle delay, one is able to attain repetition rate of one per cycle of a-c power. If 2-core configurations are used, this rate is doubled as will be demonstrated by the flip-flop illustrated. Since the output of these amplifying devices can be made many times the signal pulse power, any given stage is capable of operating many other stages as well as any required registering or indicating devices.

The condition imposed by the delay time is fortunately a very clearly definable limitation. It requires simply that the signal occur in phase with the a-c power. A pulse dictating a change in state must occur during the reset half-cycle of the amplifier and prevent that reset.

The ensuing examples will serve to illustrate these points.

Magnetic Coincidence Counter

To make a coincidence counter or AND-circuit of the single-core amplifier of Figure 1, one has only to parallel the control element with other control elements of the same type, see Figure 2. Now, during the reset half-cycle, the magnetizing current has a choice of two or more paths and will flow through that path which will provide least impedance (lowest value of E_s). If perchance one signal $E_{s,n}$ were zero, then the core would completely reset via that path and there would be zero output in the succeeding half-cycle. Only when all signals $E_{s,n}$ exist coincidentally is the flow of magnetizing current blocked and will there result an output pulse. Since this single-core configuration remains an amplifier, the indication of coincidence can occur at a power level far higher than that of the signal level. Further, the output has correct phasing and function to operate other magnetic components of similar type operating from the same power supply.

It will be noted that the output corresponding to the coincidence is not indicated immediately, but rather is delayed for the half-cycle characteristic of this type of amplification. However, coincidences can be detected at the rate of one per cycle of a-c power. Operation of two of these counters in push-pull doubles this rate.

The requirements placed upon the signals applied to the counter are that they occur throughout a major part of a reset half-cycle and that they have sufficient magnitude to block the flow of magnetizing current. The resistance R_y, shown in Figure 2 and the figures following, will be discussed at length in a following section of this paper.

Magnetic Ring Counter

A magnetic ring counter or a magnetic delay line can be constructed of single-core amplifiers as noted in Figure 3. The units are cascaded so that the output im-

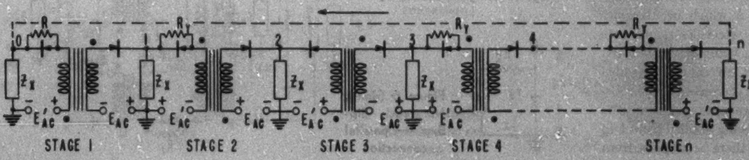


Figure 3. Magnetic delay line or ring counter

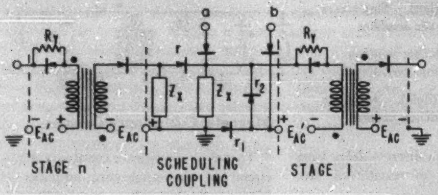


Figure 4. Scheduling coupling for ring counter

pedance of one stage is the control impedance for the following stage.

Cascading of magnetic amplifiers is normally used to provide higher gains, but the objective here is somewhat different. One desires to make use of the discrete time delay in each stage to provide a delay line.

The operation of this counter can be described in the following manner. A pulse appearing during a reset half-cycle on the input of stage 1 is reproduced on the stage 1 output the second half-cycle; this provides a signal pulse for stage 2 and so the pulse is reproduced on the stage 2 output the third half-cycle; this provides a signal pulse for stage 3 and so the pulse is reproduced on the stage 3 output the fourth half-cycle; this provides a signal pulse for stage 4 and so forth; thus the pulse is reproduced on the last stage output. Thus a pulse encounters time delay of a number of half-cycles corresponding to the number of stages n it encounters in its traverse of the delay line.

If the last stage output is the input to stage 1, the pulse would reenter the line once every n half-cycles and the line would become a ring counter. Because the pulse on the input of stage 1 must occur only on reset half-cycles, n must be an even number for ring counter operation with single-core amplifiers. Any number of stages may be used if they are made up of push-pull single-core amplifiers as noted in the previous section.

Scheduling the ring counter can be accomplished in a variety of ways. One network which has been found to be consistent with effective system grounding is shown in Figure 4. The output of stage n is not directly the input of stage 1, rather is coupled through this OR-network so that either the output of stage n or a pulse from a can provide the blocking voltage for stage 1. Erasure of a pulse from stage n or/and a can be accomplished by an inhibitor pulse from b which negates any pulse from a or n. Rectifier r_1 prevents pulse source b from being short-circuited and yet allows reset currents to flow normally when circuit connection b is absent. Rectifier r_2 prevents excessive voltage from possibly being applied from b to the reset winding if there be no pulse from a or n. Use of this scheduling system requires an isolated excitation winding for E_{ac}' of stage 1; however, all signal and output pulses may have the common ground needed for convenient coupling. Use of some type of scheduling coupling throughout a ring or delay line allows all stages to be programmed simultaneously (parallel feed) and read sequentially (series readout). This is particularly useful in telemetering wherein instantaneous information may arrive in digital form but must be read sequentially for interpretation.

The coupling impedance Z_x also may take a wide variety of forms. A simple resistance could, of course, be used here or for rectifier r_1; however, it was found advantageous to use the nonlinear impedance of Figure 5. This is a nonlinear resistor made up of a large resistance R_x, a rectifier r_x, and a source of direct voltage E_x. The voltage across this impedance is quite low when the current I_c is less than $I_x = (E_x/R_x)$. If, however, this current I_c should tend to exceed I_x, the impedance Z_x would immediately approach R_x. The magnitude of I_x is herein set greater than the sum of the magnetizing currents of the stages coupled. This means that the first core can be magnetized and the second core reset simultaneously through Z_x with practically no interaction. If, however, the first core should have been saturated, then the voltage E_{ac} would appear across the coupling impedance blocking the reset of the second core. An advantage of this coupling is that capability of the first stage to supply power to other computer or registering elements connected to this coupling remains almost undiminished. Also, voltage across the coupling due to the direct current tends to cancel the effect of voltage drops across the gating and reset rectifiers during core magnetization.

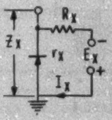
Magnetic Flip-Flop

A magnetic flip-flop or trigger circuit can be constructed of single-core amplifiers as noted in Figures 6 and 7. These are parallel- and bridge-connected versions of the same general circuit. Basically, the flip-flop is a 2-stage ring counter or a double-core ring counter stage with the output coupled to the input.

The performance of the flip-flop can be visualized by description of its response to a sequence of operations. Reference to Figure 6 will show that during any given half-cycle of a-c power, one core is being reset against whatever voltage exists across $Z_{x,1}$ and r_1. The other core is being gated and, depending upon whether or not it was reset during the previous half-cycle, may not or may allow the voltage E_{ac} to exist across $Z_{x,0}(E_0)$, and consequently through rectifier r_f, across $Z_{x,1}$. The instantaneous a-c polarities of Figure 6 indicate the condition of core I being reset and core II being gated. If core II were previously reset, there would be zero output pulse E_0 and, ruling out a pulse from a, E_1 would be zero and core I would be reset by E_{ac}'. The next half-cycle would find core I being gated with corresponding zero output pulse. Carrying this same reasoning to the following half-cycles will show the output will remain zero ("off") so long as no pulse is introduced from a.

If a pulse be now introduced from a during any half-cycle, the reset of one of the cores will be prevented and there will exist an output pulse E_0 the following half-cycle which will be introduced across $Z_{x,1}$ preventing reset of the other core and so the output will continue "on" in all following half-cycles. However, the introduction of a pulse from b during any half-cycle will cancel the pulse fed back from the load during that half-cycle and return the flip-flop to "off" until another pulse is introduced from a.

Figure 5 (below). Nonlinear interstage coupling impedance Z_x



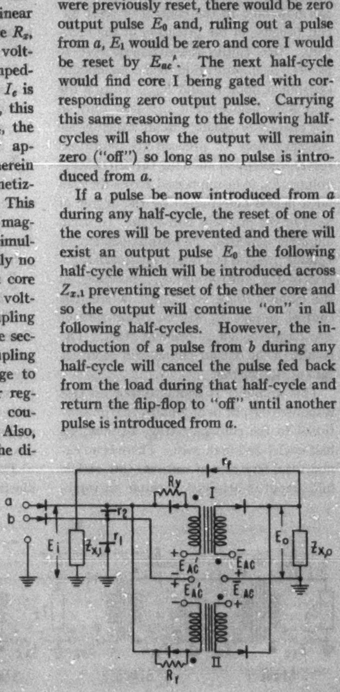


Figure 6 (right). Magnetic flip-flop — parallel connection

Generally speaking, this description of operation defines magnetic flip-flop circuits which can be switched by single pulses between two stable states. Since the magnetic device remains an amplifier in these configurations, the pulse power needed for this switching is a very small fraction of the pulse power available from the flip-flop and a small fraction indeed of the continuous power available when suitable materials are used for any given excitation frequency.

In this type of operation, the response time of the flip-flop is a half-cycle of a-c excitation frequency. This allows a maximum repetition rate of two per cycle.

The type of pulse needed for operation of this device is the same as that for the ring and coincidence counters. As a matter of fact, these devices are completely compatible with one another; that is, any type can control any of the other types as well as its own type. The signals can all be referenced to ground (as illustrated) so that complete systems can be constructed magnetically by simple connection. Since all types are amplifiers, each can control many others.

The particular flip-flop configuration illustrated cannot be efficiently pulsed "on" and "off" by sequential pulses from a single line. To accomplish this a coincidence counter is used in combination with the flip-flop and excited by the same input pulse but requiring as coincidence that the flip-flop be "on" before it will provide an "off" pulse into the flip-flop.

Experimental Illustration

Experimental units of these elements have been constructed for demonstration, see Figure 8, and have been assembled as shown in the block diagram of Figure 9. The resulting dynamic performance is shown in Figure 10.

Fragmentary computing circuits, such as shown, have little meaning except to demonstrate the elements operational characteristics. The particular configuration, Figure 9, has as objective the detection of coincidence of pulses in lines *A* and *B* (output of stages 1 and 3 of an 8-stage ring counter) and noting this coincidence by an output pulse in line *C* which triggers the flip-flop whose output appears in line *D*. The flip-flop is triggered off by any pulse in line *E* (output of stage 8 of the ring counter).

To set up the experiment, the 8-stage ring counter was first programmed via a circuit functionally the same as that of Figure 4 so that coincident pulses exist at the output of stages 1 and 3 only once per eight half-cycles. The ring counter repeats the pulse sequence as programmed every 8 half-cycles so all signals were displayed and photographed on an oscillograph screen with sweep frequency of one per eight half-cycles.

Figure 10 is this oscillogram of the voltage pulses at the various important points. It will be noted that in the second half-cycle a pulse exists at stage 1 output and a pulse exists at stage 3 output. This coincidence provides signal for the coincidence counter so it allows a pulse to exist across its output the third half-cycle. There are other noncoincident pulses from 1 and 3, but they do not excite the coincidence counter, only a coincidence is registered.

The single output pulse of the coincidence counter during the third half-cycle triggers the magnetic flip-flop on the fourth half-cycle and the output remains for four half-cycles until a pulse from stage 8 appears and triggers the flip-flop off. The "off" pulse to the flip-flop during the first half-cycle has no effect since the flip-flop is not "on."

The elements of Figure 8 were designed to operate in a system whose power frequency is 60 cycles per second and so will appear quite large as compared to units for higher frequency operation. The core size decreases rapidly as the operating frequency is increased. All cores for these designs are AECO number *4179-D1*'s with 6,615 and 4,440 turns of number 32 wire for gating and reset windings respectively. E_{ac} and E_{ac}' are nominally 50 and 35 volts respectively. R_y and R_z are 50,000 and 25,000 ohms respectively. All rectifiers are *1N57*'s. E_z is 50 volts d-c.

In the development of these units there has been no mention of limitation to specific frequencies of excitation. Theoretically, there is no limit as to the maximum excitation frequency which can be used so long as one assumes material characteristics are independent of frequency. The assumptions which were used in the development of the single-core amplifier were that the magnetic core is capable of saturation, that the residual flux approaches the saturation value, and that the rectifiers do not have leakage in excess of the magnetizing mmf of the core.

In these computing circuits the latter assumption is critical only with reference to the gating circuit rectifier. When excessive leakage exists, it exhibits its detrimental effect as a partial reset of the magnetic core during those reset periods wherein the reset circuit is blocked by an incoming signal. This partial reset prevents the complete output pulse expected the half-cycle after reset circuit blocking. It would be highly desirable to be able to compensate functionally for this leakage; that is, compensate with relative exactness at every instant in time. Fortuitously such compensation is possible both functionally and nonfunctionally. If signal pulses are made somewhat larger than E_{ac}', there is a net voltage in the reset circuit in such direction as to cause leakage in the reset circuit rectifier which would tend to compensate functionally for the gating circuit rectifier leakage. If the reset rectifier be somewhat poorer than the gating rectifier, this difference of voltage can be made proportionately smaller. Nonfunctional compensation can be accomplished quite inexpensively by degrading the reset rectifier by paralleling with the resistance R_y of Figure 2 and those following. Introducing compensating currents in other circuits, including the gating circuit, is equally effective.

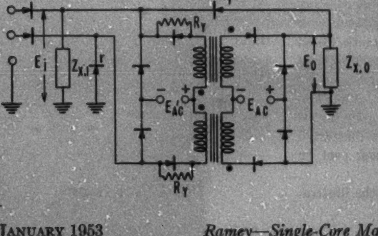


Figure 7 (left). Magnetic flip-flop — bridge connection

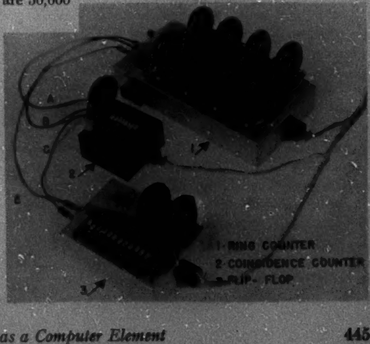


Figure 8 (right). Experimental set-up

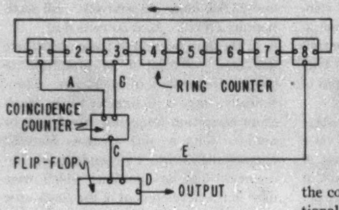


Figure 9 (left). Block diagram of experimental setup

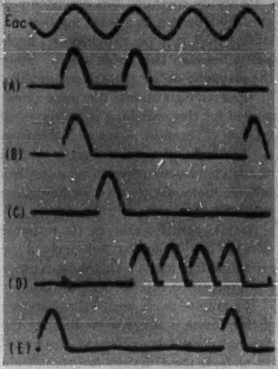


Figure 10 (right). Oscillogram of pulse voltages of interconnection lines A, B, C, D, and E of Figure 9 with E_{ac} as time reference

These and other artifices are readily imaginable for negating the effects of nonideal circuit components. The price one must pay is usually a small loss in gain for a tremendous improvement in performance. In the final analysis, however, the gain obtainable is a function of the quality of materials utilized. Aside from power gain the circuit performance is relatively independent of material quality.

There is evidence to indicate that suitable core characteristics are obtainable in the megacycle range,[7] and one can obtain crystal rectifiers for operation at these frequencies.[8] It would seem that operation of these circuits with megacycle repetition rate is possible with materials available today. The experimental work of this laboratory to date, however, has been limited to the nominal power frequency ranges useful for telemetering and remote control; the results have been highly encouraging. No fundamental difficulties are expected at higher excitation frequencies. Practical difficulties will be experienced at high frequency if relatively poor materials and components are used. For example, unless low-loss cores are used, the magnetization currents could be so high that very little gain could be expected or core heating could become a problem. There is need of more experimental work in the high-frequency ranges.

Remarks

Computer-type circuit configurations are useful in a variety of fields from telemetering and remote control to calculating machines. Though this paper has been directed to the application of the single-core magnetic amplifier to computing, it is deemed advisable to suggest its usefulness in these other fields.

Since the output of the coincidence counter is proportional to the lowest signal-voltage path for reset current, it provides an interesting sampling circuit. If one input source should monitor a quantity whose measurement is desired and its other inputs or input be pulsed periodically with a blocking pulse, the output of the coincidence counter would be proportional to the quantity being monitored. This output may also take the form of pulse-width or pulse-time modulation.

The flip-flop can be arranged to indicate and repeat the magnitude of pulses which only partially block its reset and can thereby be used for pulse stretching. In remote control systems the flip-flop can provide power for motors, indicators, and so forth.

The ring counter can be used as essentially a magnetic rotating switch for the other magnetic elements in these telemetering and control applications as it was in the computing application illustrated. As electronic computer elements are useful in these and other fields, so are their magnetic equivalents. One can easily get the impression that these circuits must be excited from a continuous source of alternating current. Actually, because of the static memory effect[9] in high-quality magnetic materials, these circuits are responsive to excitation with any sequence of alternate positive and negative pulses so long as the pulse voltage average value is reasonably constant.

The circuits, as illustrated herein, preserve the inherent gain; speed; independence of power, voltage, and frequency; wide output range; and relative independence of specific material characteristics[10] which characterize the new amplifier circuits. The speed and gain have been noted in the text. In the experimental setup, a drop in line voltage to less than 50 per cent of nominal value caused no loss of information. The output range (ratio of full or "on" output voltage to zero or "off" output voltage) can be noted by reference to the oscillogram of Figure 10 wherein the output corresponding to "on" is in the order of 100 times the output corresponding to "off." No selection of cores was necessary; standard cores were used as they came from stock with no rejects even though loop width varied by better than 2 to 1. The reliability of the circuits is essentially determined by the rectifiers used. The total power dissipation per stage is in the order of 100 milliwatts in the large stages illustrated and was localized, for the most part, in the noncritical resistance R_s.

The form of the output in the illustration is that of rectified sinusoids as dictated by the a-c line; however, almost any other form of alternating current is applicable. Square wave alternating current would give correspondingly rectangular output pulses.

The compatibility of these elements with one another and the versatility of interconnection due to the common ground is startling. In complete system design a single-type rectifier and core size can be used and the winding of all cores, regardless of the element into which they are placed, can be identical.

References

1. MAGNETIC AMPLIFIERS FOR SHIPBOARD APPLICATIONS, **L. W. Buechler.** *Electrical Engineering*, volume 68, January 1949, pages 33-37.

2. APPLICATIONS OF MAGNETIC AMPLIFIERS, **W. E. Green.** *Electronics* (New York, N. Y.), September 1947, pages 124-128.

3. MAGNETIC TRIGGERS, **A. Wang.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 38, June 1950, pages 626-29.

4. DIGITAL INFORMATION STORAGE IN THREE DIMENSIONS USING MAGNETIC CORES, **J. W. Forrester.** *Journal of Applied Physics*, American Institute of Physics (New York, N. Y.), volume 22, January 1951, pages 44-48.

5. ON THE MECHANICS OF MAGNETIC AMPLIFIER OPERATION, **Robert A. Ramey.** *AIEE Transactions*, volume 70, part II, 1951, pages 1214-23.

6. ON THE CONTROL OF MAGNETIC AMPLIFIERS, **R. A. Ramey.** *AIEE Transactions*, volume 70, part II, 1951, pages 2124-28.

7. HIGH FREQUENCY LOSS LOOPS FOR SATURABLE MAGNETIC CORES, **W. Titman, J. G. Miles.** *Report*, Engineering Research Associates, Inc. (St. Paul, Minn.), October 30, 1950.

8. SEMICONDUCTOR DIODES. *Electronics* (New York, N. Y.), volume 24, March 1951, pages 112-13.

9. STATIC MAGNETIC MEMORY, **M. Kincaid, J. M. Alden, R. B. Hanna.** *Electronics* (New York, N. Y.), volume 24, January 1951, pages 108-11.

10. THE EFFECT OF CORE MATERIALS ON MAGNETIC AMPLIFIER CIRCUITS, **J. L. Johnson.** *AIEE Transactions*, volume 71, part I, 1952, pages 26-31.

No Discussion